Refining the HIV-1 glycan shield model: dynamics of a heterogeneous envelope trimer and empirical prediction of glycan processing

Clare Garrard

GRRCLA001

Supervisor: Dr Natasha Wood

Co-supervisor: Dr Colin Anthony

Division of Computational Biology, Department of Integrative Biomedical Sciences

Presented for MScMed in Bioinformatics

Faculty of Health Sciences

University of Cape Town

11 February 2019

Declaration

1. The work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.
2. I have used the Harvard Referencing style for citation and referencing. Each contribution to and quotation in this thesis from the work(s) of other people has been cited and referenced.
3. This thesis/dissertation has been submitted to the Turnitin module and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.
4. I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate Date: 2 September 2019

Abstract

The HIV-1 surface protein, Envelope (Env), is covered in asparagine-linked glycans, which interact with the human immune system and are thus important as potential vaccine targets. Laboratory studies have shown that the glycan type and form can differ substantially at each glycan site on Env clones. However, these studies are limited by time and cost and rely on biosynthetic assumptions to elucidate the structure of branched glycans. Furthermore, glycan heterogeneity creates challenges when determining the three-dimensional structure of Env, which has resulted in the use of methods that restrict glycan processing to produce uniform glycans for these studies. Computational methods are used to complement the laboratory studies; however, due to the limitations of modelling software, even computational studies have focussed on uniformly glycosylated Env models using a limited set of high-mannose glycans, rather than a mix of glycan types.

To bridge this gap, this study set out to examine the structural differences of two computationally glycosylated HIV-1 Env trimers, one uniformly glycosylated, and the other based on the heterogeneous glycosylation of a laboratory determined gp160 strain. A secondary aim was to estimate whether the type of glycan is predictable using computational techniques, since these are less expensive and time-consuming than laboratory studies.

Using 500 ns molecular dynamics (MD) simulations, it was found that the heterogeneously glycosylated trimer had 64% greater stability, likely due to the presence of 25% more hydrogen bonds, as well as stabilising bonds which appeared to prevent asymmetrical movements. Furthermore, by focussing on the heterogeneously glycosylated trimer, a computational method based on surface area was explored to estimate the accessibility to enzymes involved in glycan processing, and to use this measure as a predictor of the glycan type.

The results of this study highlight the differences between a uniformly, and a heterogeneously, glycosylated trimer, and suggest that previous MD studies, which used uniformly glycosylated trimers, may not sufficiently describe the structural dynamics of HIV-1 Env. Notably, complex glycans appear to stabilise the trimer to a greater extent than the high-mannose glycans used in previous studies. Thus, it is evident that research on Env models should incorporate a more diverse set of glycans in order to deepen our understanding of the dynamics of Env, which will, in turn, further our understanding of its interactions with antibodies and anti-HIV compounds.

Contact: grrcla001@myuct.ac.za

Acknowledgements

The way up to the top of the mountain is always longer than you think. ~ Paulo Coelho

This has been a long road. Sometimes as I have been writing up, I have felt more like it was for a PhD than for a Masters degree. Certainly, many of the challenges have been similar: the isolation, dealing with depression, the trying to, as Dave Tabb described a PhD to me: "*go out into the bush and bring back fire*". Along the way there have been many battles, with fear and self-doubt, with improving my coding skills dramatically, with learning new programs left, right and centre, with depression, with visas, with trying to keep my priorities right (whatever right may mean in any given month…).

Fortunately, there have been fellow walkers on this long road, without whom I don't know how I would have gotten to this point.

Stephen has walked not just the Masters road with me over the last two years, but also navigated what marriage should look like. He has helped me to keep on walking, and also to remember to take a break to smell the sweet peas, pet the cats, or go dancing. Thank you.

Thanks too to my Mum, who has been a phone call away whenever I needed her – to soothe, to encourage, to offer sage words of wisdom, to laugh at the "interesting times" we live in.

Also walking this road, encouraging me, understanding when I need to dive down some other road to deal with personal things, and challenging me to look at things in new ways, has been my supervisor, Natasha. I'm grateful to have stumbled upon someone who can both push me to strive for my best, and encourage me to relax and deal with the things life has thrown my way.

Sometimes you reach a pass on the road that seems too difficult, that is designed for people more skilled than you. Some of these passes required tools that were missing from my repertoire. Fortunately, there was Dave, who took the time out to help me find the right tools so that I too could walk through the pass. I'm also hugely grateful to the HIV MD group: Natasha, Colin, Oly, Dave, and Roux-Cil.

This road would also have taken a lot longer – longer than a human lifetime – if not for the compute clusters of the Centre for High Performance Computing and the Complex Carbohydrate Research Centre. Thanks especially to Dane and Oly.

At times I found myself wondering why I was walking this road, and where it would lead next. Thank you to Dave Tabb, Shane, and Natasha, who have helped me to see the road beyond my Masters. And

thank you to Stephen for then helping to refocus my attention back on the road ahead of me, just two cats eyes ahead, so that I can finish the work that is in front of me now.

Sometimes there are road blocks that mean we have to take a detour – I'm grateful to have had Noma for company walking along the same visa detour, and to David Tabb for sharing his experiences.

Of course, where would I be without numerous cups of tea, friendly faces, wise counsel, and people who have brought light to my day? Thanks to Stephen, Mum, Rob, Sue, Ansuné, Sash, Kim, Hanna, Wes, Annie, Nikki, Philippa, and CCK in general for keeping me sane. Thank you to Kirsty for always being up for tea. Thank you to Stephen for bringing me the hot stuff with milk when I'm beyond caring what kind of tea it is! And thank you to the friendly faces in CBIO and at UCT, especially Kirsty, Ruth, Jon, Kenneth, Holy, Imane, Suresh, Chacha, Darren, Nicky, and Lizelle.

Finally, since such roads aren't freely open, I would like to thank UCT and the NRF for funding me through this degree, allowing me to fully put my mind to the question of research.

What a journey it has been… And now, gentle reader, I must not neglect to thank you too, for undertaking to read this final compilation of my Masters. I hope you will enjoy it – if not all of the pages and pages of it, at least some of the pretty pictures!

Onward!

I can do all things through him who gives me strength ~ Philippians 4:13

Contents

Declaration .....
Abstract i
Acknowledgements ii
Figures vii
Supplementary Figures vii
Tables viii
Supplementary Tables viii
Abbreviations ix
1. Introduction 1
1.1. Brief History and Discovery of the Human Immunodeficiency Virus 1
1.2. The Course of an HIV-1 Infection 2
1.3. The Structure and Lifecycle of HIV-1 3
1.4. The HIV-1 Envelope 6
1.5. Structural Studies of the HIV-1 Envelope 9
1.5.1 Crystallisation 9
1.5.2. Electron Microscopy 11
1.5.3. Mass Spectrometry and Liquid Chromatography Studies 11
1.5.4. Molecular Dynamics 12
2. Materials and Methods 15
2.1. Model Building 15
2.2. Computational Glycosylation 16
2.3. System Preparation 19
2.4. Molecular Dynamics Simulations 20
2.5. Analysis 20
2.6. Investigating EASA and SASA as Proxies for Glycan Processing 21
2.7. Graphing and Visualisation 21

2.8. Statistical Analyses 22
3. Results 23
3.1. Glycan Occupancy: Assessing the impact of glycoforms on glycan occupancy 23
3.2. RMSD: Assessing the stability of simulations 25
3.3. RMSF: Investigating the effect of different glycans on protein stability 26
3.4. Hydrogen Bonds: Investigating interactions in the glycan shield 27
3.4.1. Quantifying the glycan interactions from simulations 27
3.4.2. Assessing how glycan-glycan hydrogen bonds cluster into the defined regions of HIV-1 Env 30
3.4.3. Investigating differences in interprotomer hydrogen bond interactions 32
3.4.4. Investigating hydrogen bond longevity versus strength of bonding 36
3.4.5. Glycoform differences in hydrogen bonding 38
3.5. Changes in Solvent and Antibody Accessible Surface Area 40
3.6. Investigating Enzyme and Solvent Accessible Surface Area as a Proxy for Glycan Processing 44
3.6.1. Enzyme accessible surface area 44
3.6.2. Solvent accessible surface area 47
4. Discussion 50
4.1. The Influence of Different Glycoforms 51
4.1.1. Glycoforms influence the level of glycan occupancy due to spatial constraints 51
4.1.2. The Mixed-Trimer is more stable than the Man9-Trimer 53
4.1.3. The presence of complex glycans at the HIV-1 Env apex may be important for interprotomer stabilisation 54
4.1.4. Scissoring and trimer asymmetry in Man9-Trimer likely caused by lack of stabilising complex glycans 55
4.1.5. The 185e glycan compensated for an absent 185h glycan, highlighting the importance of interactions at the trimer apex 56
4.1.6. Hydrogen bonding is more dispersed on the Mixed-Trimer than on the Man9-Trimer 56

4.1.7. Areas around glycans, including important epitopes, exhibit the largest change in accessibility, regardless of glycoform....57
4.1.8. Repeated sites of high accessibility change may offer targets for creating glycan holes..58
4.1.9. Further glycan heterogeneity may affect the dynamics of the glycan shield....59
4.1.10. Non-identical glycosylated PNGSs may influence the dynamics....59
4.2. Potential for Glycan Processing....61
4.2.1. The choice of ER ManI as opposed to other enzymes....61
4.2.2. The impact of oligomerisation....61
4.2.3. EASA may be useful for differentiating high mannose and complex glycans....61
4.2.4. Higher predictive power of the EASA compared to the SASA....62
4.2.5. EASA indicates potential for a higher number of complex glycans in the Mixed-Trimer 62
4.2.6. Glycoforms chosen at variable and uncharacterised sites align well with EASA predictions 62
4.2.7. The effect of rotamers on accessibility may not be sufficiently represented....63
4.2.8. Processing prediction and glycan holes....63
5. Conclusion....64
Appendix....66
Supplementary Methods: Algorithms written for this project....72
References....78

Figures

Figure 1: Changes in disease markers and immune response over the course of an HIV-1 infection...3
Figure 2: Structure of HIV-1 and extensive heterogeneity of its glycan shield.....4
Figure 3: Structure of the HIV-1 Envelope.....6
Figure 4: Simplified overview of glycans used and their degree of processing 8
Figure 5: Differential glycosylation of trimers 17
Figure 6: Percentage glycan occupation at each PNGS across ten models 24
Figure 7: Graph to compare RMSD of the Mixed-Trimer and Man9-Trimer 25
Figure 8: Comparison of RMSF for Mixed-Trimer and Man9-Trimer 26
Figure 9: Comparison of inter-glycan hydrogen bonding for the Man9-Trimer and Mixed-Trimer...29
Figure 10: Comparing glycan-glycan hydrogen bonds by HIV-1 region 31
Figure 11: Comparison of interprotomer hydrogen bonding of the Man9-Trimer and Mixed-Trimer34
Figure 12: Scissoring of the Man9-Trimer 35
Figure 13: UpSet plot of common glycan sites involved in dominant interprotomer interactions.....36
Figure 14: Location of predominantly present and absent glycans in hydrogen bonding 38
Figure 15: Surfaces for antibody and solvent accessible surface area calculations.....40
Figure 16: Solvent accessible and antibody accessible surface area of protein residues.....41
Figure 17: Sites of highest change in AbASA on the Man9-Trimer and Mixed-Trimer.....42
Figure 18: Sites of highest SASA change on the Man9-Trimer and Mixed-Trimer 42
Figure 19: Basis of enzyme accessible surface area calculations for investigating processing prediction 45
Figure 20: Investigating enzyme accessible surface area for processing prediction 46
Figure 21: Comparison of EASA prediction with Mixed-Trimer glycoforoms 47
Figure 22: Investigating solvent accessible surface area as a proxy for glycan processing prediction 48
Figure 23: Comparison of processing levels of SASA-predicted Trimer, EASA-predicted Trimer, and Mixed-Trimer.....49

Supplementary Figures

Supplementary Figure 1: HXB2 aligned 4TVP template with PNGSs highlighted 66
Supplementary Figure 2: Heatmap showing glycan-glycan hydrogen bonding on Man9-Trimer 67
Supplementary Figure 3: Heatmap showing glycan-glycan hydrogen bonding on Mixed-Trimer 68

Tables

Table 1: Glycans used for computational glycosylation, in abbreviated and expanded forms.....16
Table 2: Glycoforms at each PNGS on the Man9-Trimer and Mixed-Trimer 19
Table 3: Potential N-glycosylation sites that were not glycosylated on the final models 25
Table 4: Potential for hydrogen bonds.....28
Table 5: Dominant interprotomer interactions for the Man9-Trimer and Mixed-Trimer (≥5000).....33
Table 6: Glycans that were predominantly present or absent in hydrogen bonding.....37
Table 7: Comparison of sites where glycoforms differ 39
Table 8: Amino acids featured in the top 10 sites of SASA and AbASA change 43

Supplementary Tables

Supplementary Table 1: Top 10 Ranked RMSF values used for delineating outliers.....69
Supplementary Table 2: Ranking of glycan percentage contribution to hydrogen bonding 70

Abbreviations

Ab: Antibody

AbASA: Antibody accessible surface area

AIDS: Acquired immunodeficiency syndrome

ARVs: Antiretroviral drugs

bNAb: Broadly neutralising antibody

C1 – C5: Constant regions 1 – 5 of gp120

Complex-glycan: DNeup5Aca2-6DGalpb1-4DGlcpNAcb1-2DManpa1-6[DNeup5Aca2-6DGalpb1-4DGlcpNAcb1-2DManpa1-3]DManpb1-4DGlcpNAcb1-4DGlcpNAcb1-OME
NB: *"Complex-glycan" is used for this specific glycan, while "complex glycan" describes the level of glycan processing*

CFG notation: Consortium for Functional Genomics notation

EASA: Enzyme accessible surface area

EM: Electron microscopy

Env: Envelope protein

ER: Endoplasmic reticulum

ER Man I: Endoplasmic reticulum alpha-mannosidase I

GlcNAc: N-acetylglucosamine

gp120: Glycoprotein 120 subunit of HIV-1

gp41: Glycoprotein 41 subunit of HIV-1

HIV-1: Human Immunodeficiency Virus type 1

HILIC-UPLC: Ultra-performance liquid chromatography

HAART: Highly active anti-retroviral therapy

IMP: Intrinsic mannose patch

LC:	Liquid chromatography
Man:	Mannose
Man9:	Mannose 9; $Man_9GlcNAc_2$; DManpa1-2DManpa1-6[DManpa1-2DManpa1-3]DManpa1-6[DManpa12DManpa1-2DManpa1-3]DManpb1-4DGlcpNAcb1-4DGlcpNAcb1-OH
Man9-Trimer:	HIV-1 Env trimer glycosylated with Man9 glycans
Man5:	Mannose 5; $Man_5GlcNAc_2$; DManpa16[DManpa1-3]DManpa1-6[DManpa1-3]DManpb1-4DGlcpNAcb1-4DGlcpNAcb1-OH
MD:	Molecular Dynamics
Mixed-Trimer:	HIV-1 Env trimer glycosylated with a mix of Man9, Man5 and Complex-glycans
MS:	Mass spectrometry
NAbs:	Neutralising antibodies
PNGS:	Potential N-linked Glycosylation Site (Asparagine-X-Threonine/Serine motifs, where X ≠ Proline)
RMSD:	Root mean squared deviation
RMSF:	Root mean squared fluctuation
SASA:	Solvent accessible surface area
TAMP:	Trimer-associated mannose patch
V1-V5:	Variable regions 1-5 of HIV-1 Env

1. Introduction

If I have seen further, it is only by standing on the shoulders of giants ~ Isaac Newton

1.1. Brief History and Discovery of the Human Immunodeficiency Virus

In 1983, the causal agent of Acquired Immunodeficiency Syndrome (AIDS) was identified as the Human Immunodeficiency Virus (HIV) by two independent laboratories (Barré-Sinoussi *et al.*, 1983; Gallo *et al.*, 1983). The virus was found to primarily target CD4+ T cells, depleting them over time to cause the immune deficiencies that had been observed (Klatzmann *et al.*, 1984). Due to its high genetic variability, a cure has proven elusive, and HIV has gone on to become a global pandemic that has caused more than 35 million deaths (WHO, 2016).

HIV has two types – HIV-1, and HIV-2 – and it is the more virulent HIV-1 which is responsible for the current global pandemic. There are 4 HIV-1 groups comprising lineages M (main), N (non-M or O), O (outlier) and, most recently identified, group P (Buonaguro, Tornesello and Buonaguro, 2007; Plantier *et al.*, 2009). Each of these arose from independent cross-species transmission events from chimpanzees or gorillas to humans (Sharp and Hahn, 2011). The main group, M, is responsible for the majority of HIV-1 cases worldwide, which numbered approximately 36.9 million at the end of 2017 (WHO, 2017). It is made up of 9 clades/subtypes: A, B, C, D, E, F, G, H, J and K, as well as recombinant forms that result in a mix of clades. Clades A, B, and C are the most common forms of HIV-1: Clade A is found mostly in East Africa, Central Asia, and Eastern Europe; Clade B in America, Western Europe and Australia; and Clade C in Southern Africa, the Horn of Africa, and India (McCutchan, 2006). Clade C alone accounts for more than 50% of HIV-1 cases (Günthard and Scherrer, 2016; Shrivastava *et al.*, 2018).

Given that the majority of HIV-1 cases occur in the developing world, problems of access to HIV-1 interventions such as anti-retrovirals (ARVs) and highly active anti-retroviral therapy (HAART) persist (Atun *et al.*, 2016; Stephenson and Barouch, 2016). While these interventions have changed the outlook of HIV-1 from being a lethal disease to a manageable condition, only 2 in 5 people have access to ARVs (Stephenson and Barouch, 2016), highlighting the need for a vaccine to help bring the HIV-1 pandemic under control. Most vaccines are created using neutralising antibodies (NAbs) to guide vaccine design (Kwong, Mascola and Nabel, 2011), and thus an important step towards a vaccine

for HIV-1 has been understanding the immune responses of the body once HIV-1 infection has occurred.

1.2. The Course of an HIV-1 Infection

The majority of HIV-1 infections are caused by a single founder virus, passed across the mucosal membrane of an HIV-positive person to an HIV-negative person (Keele *et al.*, 2008). There is a period of about 10 days where viral DNA cannot be detected in the plasma, known as the eclipse phase (McMichael *et al.*, 2010). Thereafter, viral levels increase exponentially, mediated by dendritic cell dissemination of HIV-1 to the lymphoid tissue (Geijtenbeek *et al.*, 2000). The gut-associated lymphoid tissue (GALT) is particularly affected, and a vast number of CD4+ T cells are lost from this immune cell rich region during acute infection (Figure 1) (Brenchley *et al.*, 2004).

As the viral load increases, naïve B cells are exposed to HIV-1, and the first antigen-specific B-cells appear, although their expressed antibodies are not yet capable of neutralising the virus. Within 12 weeks of infection, the body starts to produce NAbs against HIV-1 (Figure 1) (Richman *et al.*, 2003). Thereafter, 3 – 6 months post-infection, the viral load attains a stable level, known as the viral set point (McMichael *et al.*, 2010), and CD4+ T cell levels rise again with the decrease in the viral load. This chronic phase can last for years, especially with treatment. Without treatment, an HIV-1 infection will usually progress, leading to immune suppression, AIDS-related illnesses, and finally death as CD4+ T cells are destroyed.

In 10 – 30% of HIV-positive individuals, broadly neutralising antibodies (bNAbs) are produced after 2 to 3 years of infection (Figure 1) (Sather *et al.*, 2009; Walker *et al.*, 2010; Gray *et al.*, 2011). Unlike the NAbs produced by the 12th week of infection, these antibodies are capable of neutralising more than one strain of the virus. Given that the founder virus rapidly evolves to form a swarm of millions of variant viruses known as a "quasi-species" (Figure 1) (Eigen and Schuster, 1977; Domingo *et al.*, 1978), this broad neutralising ability is of great interest to vaccine research. By exploiting the co-evolution of the virus and the antibody response (Liao *et al.*, 2013), researchers hope to use rational HIV-1 immunogen design in order to trigger the formation of bNAbs.

There have been numerous vaccine trials (Klein *et al.*, 2013), but the most effective HIV-1 vaccine had only a moderate efficacy of 31.2% (Rerks-Ngarm *et al.*, 2009). It is clear that better immunisation strategies are needed, and a fuller understanding of the structure of HIV-1 and its relation to bNAbs has emerged as a vital step towards this goal.

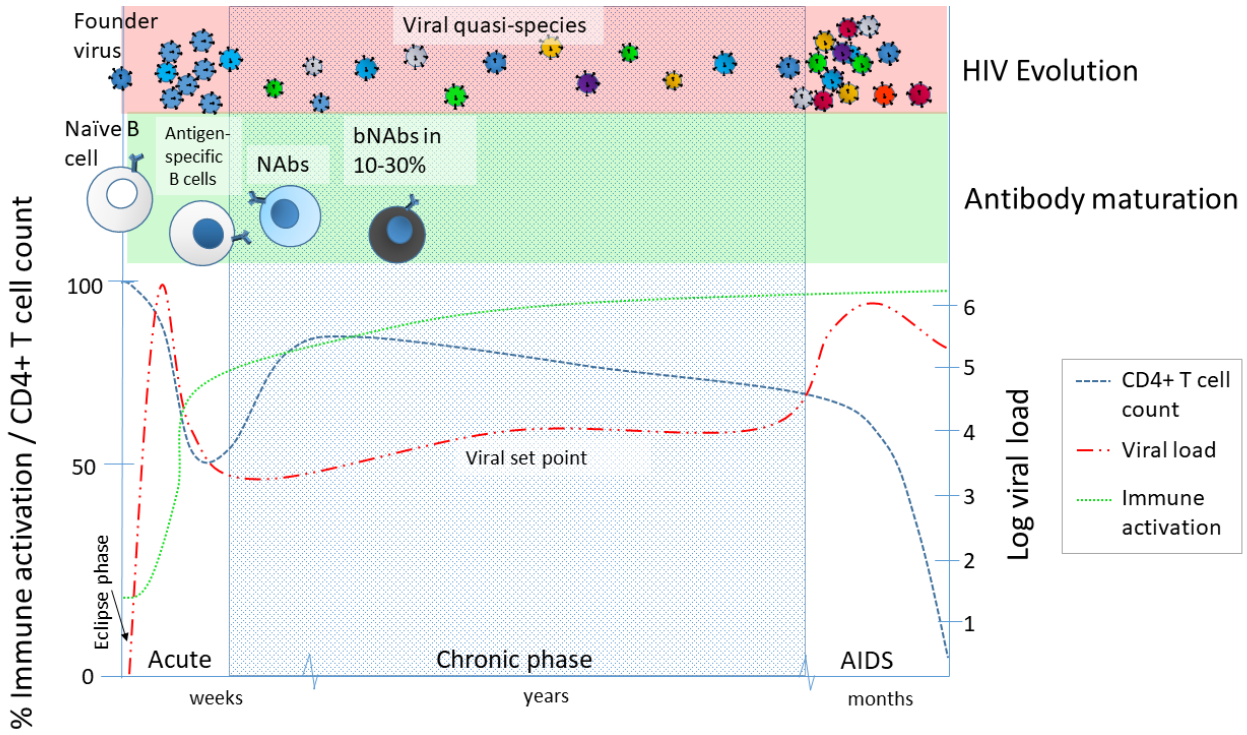


Figure 1: Changes in disease markers and immune response over the course of an HIV-1 infection

Measures of disease progression, along with stages in HIV evolution and antibody maturation over the course of an HIV-1 infection. Adapted from Grossman *et al*. 2006.
HIV Evolution: A single founder virus begins the HIV-1 infection, and evolves to form a quasi-species made up of millions of variants. Antibody maturation: Naïve B cells are exposed to HIV-1, resulting initially in the production antibody specific B cells. Within 12 weeks NAbs are produced. In 10 – 30% of individuals, bNAbs are produced within 2-3 years of infection. Eclipse phase: short ~10 day period after infection when the virus cannot be detected. Acute phase: a period of weeks where viral load increases rapidly, decimating the CD4+ T cell population until it is controlled by an increase in immune activation, and settles at the viral set point. Chronic phase: a period of years of interplay between immune activation and HIV-1 immune evasion, with a steady decrease in CD4+ T cell count. AIDS: Once the CD4+ T cell count reaches a critical level (<200), the levels of viraemia can no longer be controlled. The increased viral load depletes the remaining CD4+ T cells, and opportunistic infections set in, leading to death.
** All images are my own unless otherwise stated*

1.3. The Structure and Lifecycle of HIV-1

HIV-1 is an enveloped retrovirus of the *Lentivirus* genus. It has an average diameter of 145nm and is made up of 9 genes (Figure 2A) (Briggs *et al.*, 2003). Encased inside the host-derived lipid membrane, within the viral matrix, is the conically shaped capsid protein (Figure 2A). This contains viral enzymes, accessory proteins, and two copies of single stranded genomic RNA tightly bound to zinc-finger nucleocapsid proteins (Summers *et al.*, 1992; Frankel and Young, 1998). On the surface of the virion are 5 – 15 envelope proteins (Env), which extend outside the lipid membrane (Figure 2A and B) (Brandenberg *et al.*, 2015), and are covered in a huge variety of N-linked glycans.

Glycans are grouped into three broad classes: high mannose, where only mannose (Man) residues extend the $Man_3GlcNAc_2$ core (Figure 2C); hybrid glycans, where one arm is extended by mannose

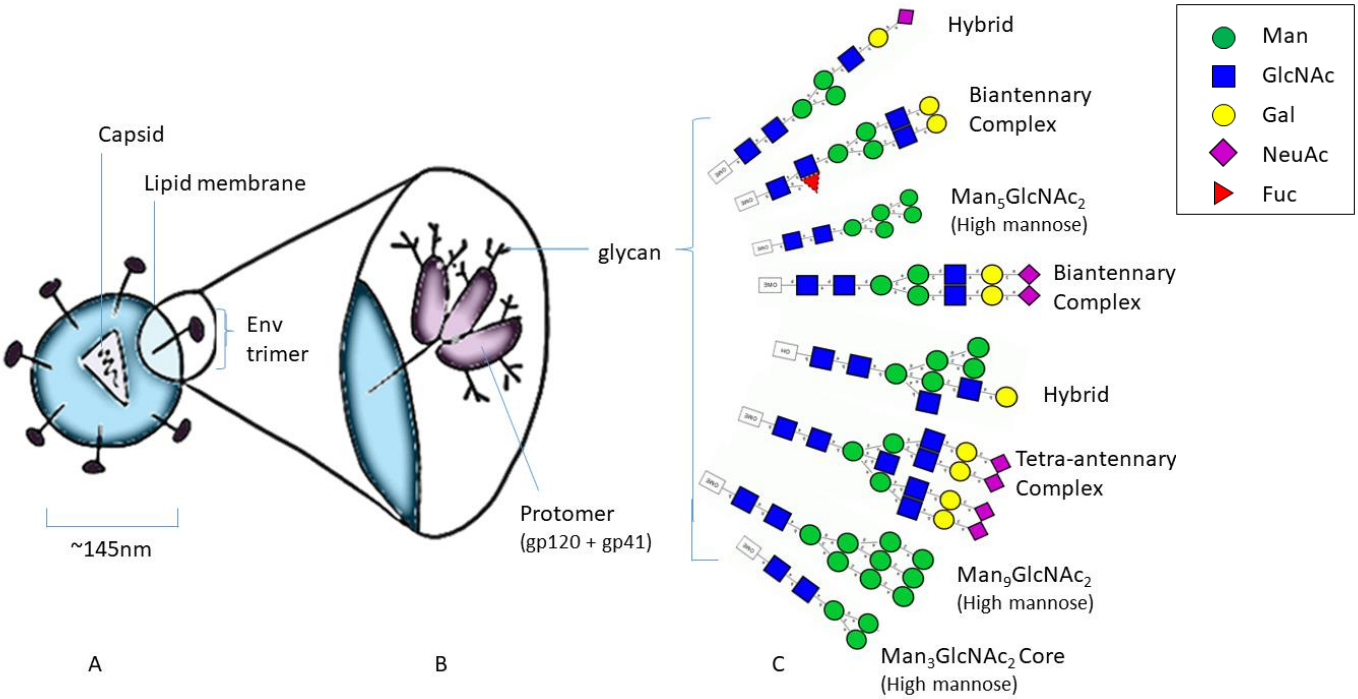


Figure 2: Structure of HIV-1 and extensive heterogeneity of its glycan shield

A: Stylised drawing of HIV-1. B: Magnification to illustrate the location of glycans on the HIV-1 Envelope trimer. C: A small subset of the vast glycan diversity in the glycan shield, shown in Consortium for Functional Genomics (CFG) Notation. The broad categories of glycan are indicated (high mannose, hybrid and complex), and the $Man_9GlcNAc_2$ and $Man_5GlcNAc_2$ glycans are shown, along with the $Man_3GlcNAc_2$ glycan which forms the core of all N-linked glycans.

residues, and the other by the addition of N-acetylglucosamine (GlcNAc) residues; and complex glycans, which are extended by arms, or "antennae", each connected by a GlcNAc residue to the core. Complex glycans are commonly described along with their number of antennae (e.g. the biantennary complex glycans in Figure 2C), and can have up to six branches (Stanley, Taniguchi and Aebi, 2017).

The Env trimer spikes are crucial for HIV-1 infectivity, since they interact with human CD4+ T cells during the processes of binding to the CD4 receptor and a CC or CXC chemokine coreceptor, fusing the host and viral membranes, and facilitating entry of the viral capsid into the newly infected immune cell (Engelman and Cherepanov, 2012). Because these interactions require conformational changes in Env, it is a metastable complex (Julien, Cupo, *et al.*, 2013). The prefusion state of Env is present on circulating virions; upon receptor binding, Env is activated, and undergoes conformational changes that result in membrane fusion and the entry of the viral core into the host cell (Chan and Kim, 1998).

Once HIV has entered the cell, viral reverse transcriptase (RT) creates a double-stranded copy of viral DNA from the two ~9 kilobase RNA strands in a low fidelity copying process, with an average error rate of 1 error for every 1700 nucleotides (Roberts, Bebenek and Kunkel, 1988). This high error rate, along with the presence of 2 RNA strands, which allows for recombination to occur, contributes to the

genetic diversity of HIV-1 (Freed, 2015). The newly formed DNA is transported across the nuclear membrane to enter the host nucleus. Here, it is integrated into the host genome by viral integrase in order to carry out replication (Engelman and Cherepanov, 2012).

At this point, if the CD4+ T cell enters a resting memory state, rather than replicating, the virus will stay embedded in the genome as a latent viral reservoir (Siliciano and Greene, 2011). Otherwise, viral replication takes place, using the host machinery to transcribe a copy of viral RNA with the help of the viral Tat protein (Engelman and Cherepanov, 2012). The HIV-1 RNAs are transported to the cytoplasm, where they use the host machinery to produce viral proteins (Freed, 2015).

The HIV-1 genome is translated into 3 major polyproteins: Gag, Gag-Pol, and Env. These are then broken down to form individual proteins. The 55 kDa Gag polyprotein forms the major internal structural proteins: the matrix, capsid and nucleocapsid, as well as the budding protein, p6 (Freed, 2015). The 160 kDa Gag-Pol polyprotein precursor contains the viral enzymes: reverse transcriptase, protease and integrase (Freed, 2015). The Env polyprotein is modified during the process of protein translation with sugars, known as glycans, being added in the process of glycosylation (Figure 2B and C) (Checkley, Luttge and Freed, 2011). The glycans make up roughly half of the molecular weight of Env (Lasky *et al.,* 1986). The 160 kDa gp160 Env polyprotein is cleaved by furin to form the gp120 and gp41 subunits (Checkley, Luttge and Freed, 2011). These subunits remain non-covalently associated as a protomer; three covalently-bound protomers make an HIV-1 Env trimer spike (Figure 2B) (Checkley, Luttge and Freed, 2011).

The Gag and Gag-Pol polyproteins, along with two copies of genomic RNA per virion, assemble at the cell membrane, where Gag recruits lipid rafts made up of cholesterol- and glycophingolipid-enhanced microdomains. The lipid rafts are believed to be important for viral assembly. Glycosylated Env trimers also assemble at the host cell membrane, with their gp41 cytoplasmic tails embedded within the membrane, and the gp120 components extending outside the membrane. The model of incorporation of Env into new immature viruses is not entirely clear, but evidence points to the interaction of the gp41 cytoplasmic tail of Env and the matrix constituent of Gag during the incorporation process (Freed, 2015).

The p6 domain of the Gag polyprotein hijacks the cellular endosomal sorting complex required for transport (ESCRT) in order to mediate the budding off of a new virion by pinching the host cell membrane together. During or soon after viral release, the viral protease cleaves sites in the Gag and Gag-Pol polyproteins, triggering viral maturation (Freed, 2015). The new viruses then go on to infect other immune cells.

Having budded off from the host membrane, the Gag and Gag-Pol viral proteins are contained within the host-derived viral membrane, and the surface protein, Env is the only exposed HIV-1 antigen that can be targeted by neutralising antibodies (Wyatt and Sodroski, 1998; Zhou *et al.*, 2007). Env thus forms the basis of much research, as it represents an important vaccine target.

1.4. The HIV-1 Envelope

The trimeric Env protein is made up of 3 covalently associated gp120-gp41 protomers (Liu *et al.*, 2008). The gp120 subunit has five variable regions (V1-5) separated by disulphide bonds into flexible loops. These are interspersed with constant regions (C1-5) (Figure 3 A – C) (Starcich *et al.*, 1986;

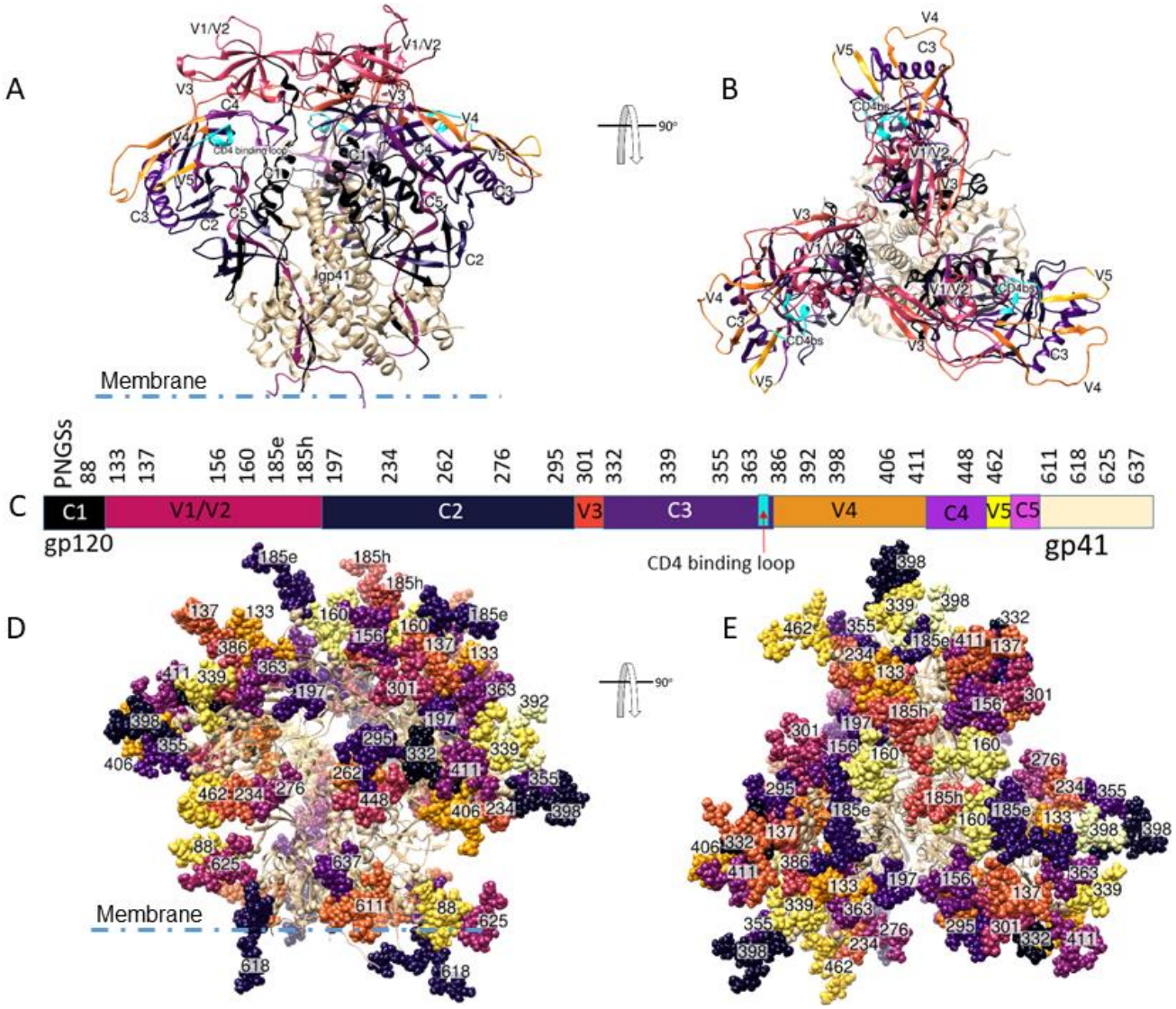


Figure 3: Structure of the HIV-1 Envelope

A and B: (A) Side view and (B) Top view of HIV-1 Env protein showing gp41 ectodomain, and variable and constant regions of gp120. Colours of each region correspond to those in C. C: Linear schematic of HIV-1 Env regions and the PNGSs within each. Not to scale. D and E: (D) Side view and (E) Top view of glycosylated HIV-1 Env protein. Glycans are numbered according to the HXB2 reference sequence (See Supplementary Figure 1). Colours here were used to create the maximum distinction between neighbouring glycans, and do not relate to the regions presented in C.

Willey *et al.*, 1986; Leonard *et al.*, 1990). The constant regions are important for binding to the CD4 receptor when infecting CD4+ T cells. This binding occurs in a recessed pocket within gp120, starting off by early contact with the CD4-binding loop (Figure 3A and B) (Zhou *et al.*, 2010; Wu *et al.*, 2011). The variable regions on the other hand play an important role in immune evasion, as their length and sequence change over the course of an infection, creating a challenge for circulating antibodies (Curlin *et al.*, 2010). The V1/V2 and V3 loops make up the trimer apex, while the V4 and V5 loops project outwards from the gp120 core (Figure 3A and B) (Ward and Wilson, 2015). The gp41 subunit is made up of three major domains: an extracellular ectodomain, a transmembrane domain, and a cytoplasmic tail (Checkley, Luttge and Freed, 2011).

Each gp120-gp41 protomer has 25 to 30 highly polar N-linked glycans, totalling ~90 glycans if all potential sites are glycosylated (Figure 3C – E) [1] (Kong, Stanfield and Wilson, 2013; Cao *et al.*, 2017). Importantly, glycans are the first point of contact with the human immune system – affecting which cells HIV-1 will enter according to the CC or CXC chemokine co-receptors (Wood *et al.*, 2013), shielding protein epitopes from antibodies (Wei *et al.*, 2003), and paradoxically also forming targets for broadly neutralising antibodies (Walker *et al.*, 2010; Gray *et al.*, 2011). This makes them important as vaccine and therapeutic targets.

The increased glycosylation of viral sequences over the course of an infection highlights the importance of glycans to immune evasion (Sagar *et al.*, 2006; Curlin *et al.*, 2010). This trait has been successfully exploited by creating a "glycan hole" through the removal of 4 glycans in order to prime powerful neutralising antibodies (Zhou *et al.*, 2017). A natural hole seen on the BG505 SOSIP.664 construct was also found to elicit potently neutralising monoclonal antibodies (McCoy *et al.*, 2016). Interestingly, removal of a single glycan does not always result in a hole, but instead can result in structural rearrangements of the glycan shield (Ferreira *et al.*, 2018). Similarly, glycans have been used to mask unimportant antibody targets on the Env trimer in order to redirect the immune response towards important epitopes (Duan *et al.*, 2018). These approaches may be helpful in the eventual design of a vaccine.

The process whereby glycans are attached to the protein as it is made is known as glycosylation. Glycosylation of HIV-1 Env is carried out by host enzymes found in the endoplasmic reticulum (ER), and Golgi apparatus (Doores, 2015). In the ER, the precursor sugar, $Glc_3Man_9GlcNAc_2$, is added to the Env polyprotein at asparagine residues found within special motifs, known as potential N-linked

[1] Glycans are referred to using the HXB2 numbering system (See Supplementary Figure 1 for sequence alignment). This convention is used throughout this work.

glycosylation sites (PNGSs). PNGSs have the amino acid sequence Asparagine-X-Threonine/Serine (or NXT/S), where X can be any amino acid except proline. (Kornfeld and Kornfeld, 1985). Once attached to Env at PNGSs, the precursor sugars are processed to form $Man_9GlcNAc_2$ (henceforth abbreviated to Man9) (Moremen, Tiemeyer and Nairn, 2012).

The barrel-shaped ER α-mannosidase I enzyme (ER Man I) removes one mannose residue from Man9, to form the $Man_8GlcNAc_2$ glycan. After this, the glycosylated gp160 Env polyproteins group together (usually in sets of three) in the process of oligomerisation, and move to the Golgi apparatus. (Earl, Doms and Moss, 1990; Moremen, Tiemeyer and Nairn, 2012; Xiang, Karaveg and Moremen, 2016).

In the Golgi, the glycans are trimmed to form $Man_5GlcNAc_2$ (abbreviated to Man5) by Golgi α-mannosidases IA – C (Doores *et al.*, 2010). The α-mannosidase III enzyme removes the two outermost mannose residues to form $Man_3GlcNAc_2$, which is the core structure of all N-linked glycans (Stanley, Taniguchi and Aebi, 2017). Sugar residues are added one by one, elongating the glycans, which progress through a hybrid stage to form complex glycans (Figure 4).

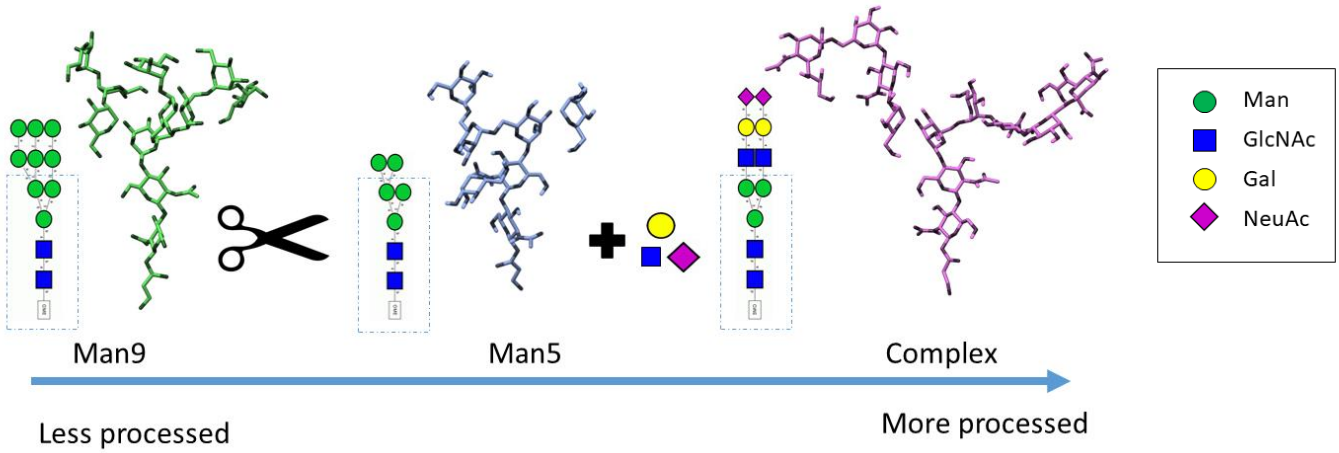


Figure 4: Simplified overview of glycans used and their degree of processing

Man9 is the first glycan in the glycosylation assembly line. It is trimmed, residue by residue, to form Man5. As other sugar residues are added, the glycan progresses through a hybrid stage before complex glycans are formed. Glycans are depicted in order from less to more processed with CFG notation (left) and structural visualisation (right) shown for each glycan. Blue boxes are used to highlight the $Man_3GlcNAc_2$ core of each glycan. The set of glycans shown is important to this work.

The availability of hexosamine is an important factor in determining the level of branching of hybrid and complex glycans (Lau *et al.*, 2007). Glycans with two branches are referred to as bi-antennary (such as the Complex-glycan in Figure 4), those with three are referred to as tri-antennary, and those with four branches are referred to as tetra-antennary.

Where processing enzymes cannot fit due to spatial constraints, glycans remain underprocessed, resulting in the characteristic high mannose areas found on HIV-1 – the intrinsic mannose patch (IMP) on the outside of the gp120 monomers (Doores *et al.*, 2010), as well as the trimer-associated mannose patch (TAMP) in the centre of the trimer (Pritchard, Vasiljevic, *et al.*, 2015). These mannose patches are conserved across strains and over the course of an HIV-1 infection (Coss *et al.*, 2016). The relationship between solvent accessible surface area and glycan processing has recently been investigated on a large set of glycoproteins from the Protein Data Bank, highlighting the link between spatial constraints and lower levels of glycan processing (Suga, Nagae and Yamaguchi, 2018).

The characteristic high mannose regions have been shown to be important to the elicitation of bNAb responses (Lavine *et al.*, 2012), and studies using kifunensine to restrict glycan processing beyond Man9 have been found to improve the neutralising ability of V3 Abs for JR-FL Env immunogens, and V2 and V3 Abs for BaL viral Env immunogens (Upadhyay *et al.*, 2014).

Importantly, although the glycosylation process takes place in an ordered assembly line, factors such as spatial constraints, substrate availability, speed of translocation, cell type, rotation about glyosidic linkages, viral strain, amino acid divergence, and enzyme compartmentalisation result in extensive glycan heterogeneity (Figure 2C) (Varki *et al.*, 2009; Pritchard, Harvey, *et al.*, 2015). Thus, after processing, a Man9 glycan may result in any of multiple $Man_{5-8}GlcNAc$ forms (depending on which residues are removed in turn); it may be reduced to the $Man_3GlcNAc_2$ core and not further extended; it could result in a bi-antennary hybrid glycan, or a hexa-antennary complex glycan, or any of multifarious glycan forms – not all of which are yet known. This heterogeneity manifests within and across sites on any single Env, as well as between the different Envs on the same viral surface, within the viral isolate and between different strains (Doores, 2015; Behrens *et al.*, 2016; Panico *et al.*, 2016; Cao *et al.*, 2017, 2018).

1.5. Structural Studies of the HIV-1 Envelope

1.5.1 Crystallisation

Glycan heterogeneity has created special challenges for structural studies of Env. Crystallisation is made difficult by the presence of different glycoforms and glycan conformations, since these disrupt the uniformity that is required for the formation of crystals (Chang *et al.*, 2007). For decades, X-ray crystallography was limited to proteins with low levels of glycosylation; this was commonly accepted as not exceeding 1 PNGS per 100 amino acids (Gristick, Wang and Bjorkman, 2017). HIV-1 Env by contrast has roughly 1 PNGS for every 20 amino acids, and was thus initially studied through removing PNGSs, using high mannose glycosylation systems, deglycosylating the proteins, or a combination of

these approaches (Gristick, Wang and Bjorkman, 2017). Indeed, the first crystal structures of gp120 were achieved using insect cell expression systems to result in fewer glycans of a high or pauci-mannose type, and then using endoglycosidases EndoH and EndoD to remove glycans prior to crystallisation (Kwong *et al.*, 1998; Wyatt *et al.*, 1998; Kwong, 1999).

As the importance of glycans has become more apparent by their presence in bNAb epitopes, it has become common to attempt to preserve more of the glycan shield in experimental studies. For crystal studies, homogeneity is still desirable, and thus glycan processing has been constrained in order to produce ordered crystals. Kifunensine is used to prevent further processing from the Man9 starting glycan; GnTI-deficient cells are used to stop processing from Man5; and swainsonine is used to create a shield of hybrid-type glycans (Chang *et al.*, 2007).

Along with these methods of inhibiting glycan processing to maintain uniformity, antibodies have also been used to shield the glycans they are complexed with, along with surrounding glycans (Julien, Cupo, *et al.*, 2013). This allows insight into the specific glycoforms that are present in the bound and shielded area. The discovery of numerous and diverse bNAbs since 2009 (Montefiori, 2009; Scheid *et al.*, 2009; Simek *et al.*, 2009; Walker *et al.*, 2009) has drastically improved crystal screening, allowing increasingly complete crystal structures to be resolved. This played an important part in obtaining HIV-1 Env trimer crystal structures.

Achieving native trimeric crystal structures took many years; hindered by the low yield of viral spikes and the metastable nature of the Env glycoprotein, researchers turned to recombinant forms. This started in the early 2000s, with the initial forms of the JR-FL SOSIP trimer construct (Binley *et al.*, 2000; Sanders *et al.*, 2002), but it was not until 2013 that the critical breakthrough came in the form of the clade A BG505-SOSIP.664 recombinant trimer (Sanders *et al.*, 2013). By adding disulphide bonds between the gp120 and gp41 subunits (SOS) to stop their dissociation, substituting an isoleucine in gp41 with a proline (IP) to prevent fusion-induced changes, and truncating the glycoprotein at position 664 to include only the region of Env outside the plasma membrane, the Moore lab was able to crystallise the trimer in its prefusion state (Sanders *et al.*, 2013). This recombinant represents one of the leading immunogen prospects for an HIV-1 vaccine (Doores, 2015; Sanders *et al.*, 2016).

In 2016, fully glycosylated trimers, that is, where the free glycans were not removed by endoglycosidases prior to crystallisation, were crystallised by the Kwong lab. Using GnTI-deficient cell lines, which resulted in high mannose trimers of glycoforms $Man_{5\text{-}9}GlcNAc_2$, along with antibodies PGT122, 35O22, and VRC01, they crystallised trimers from clades A (BG505.664), B, and G,

managing to preserve ~50% of the glycan mass (Stewart-Jones *et al.*, 2016). This represented a great advance in visualising the glycan shield of HIV-1 Env.

1.5.2. Electron Microscopy

While crystallisation is viewed as the "gold standard" of structural biology, it is important not to overlook the enormous contribution of electron microscopy (EM) when it comes to elucidating glycoprotein structure. In contrast to crystallisation, EM is enhanced by glycans, as they provide more mass and features, allowing computer algorithms greater fidelity in the identification, sorting, and aligning processes (Crispin, Ward and Wilson, 2018). Notably EM played a vital role in developing the BG505 SOSIP.664 immunogen (Sanders *et al.*, 2013).

1.5.3. Mass Spectrometry and Liquid Chromatography Studies

Crystallisation and electron microscopy have been very useful in revealing the HIV-1 Env protein structure and glycan positions; however these methods represent a form of averaging when it comes to the chemical composition of the glycans themselves (Crispin, Ward and Wilson, 2018). Given the importance of glycans to bNAb epitopes, there has been an increasing interest in more precisely determining the chemical makeup, and variability in this makeup (or microheterogeneity) of glycans at each PNGS on HIV-1 Env using mass spectrometry (MS) and liquid chromatography (LC).

The first study was carried out nearly 30 years ago on recombinant monomeric gp120 using reverse-phase high-performance liquid chromatography (RP-HPLC) and pointed to a high mannose dominated glycan shield (Leonard *et al.*, 1990); since then mass spectrometry advances have allowed more detailed insight into the microheterogeneity and occupancy levels of glycans at each PNGS. Studies were initially restricted to recombinant monomeric gp120 constructs due to the challenges of isolating native virions. Soon the comparison of recombinant gp120 with virus associated gp120 facilitated the characterisation of the intrinsic mannose patch, which persists across strains and is distinct from host-cell processing (Doores *et al.*, 2010). The eventual move to using recombinant trimers in mass spectrometry hinted at a difference in processing from a gp120 monomer (Guttman *et al.*, 2014); this was fully clarified in a in a hydrophilic interaction ultra-performance liquid chromatography (HILIC-UPLC) study with the identification of the trimer-associated mannose patch (Pritchard, Vasiljevic, *et al.*, 2015)

Recent mass spectrometry studies (Behrens *et al.*, 2016; Panico *et al.*, 2016; Cao *et al.*, 2017, 2018) focus on the glycan specificity at each glycosylation site on trimeric Env. Two main approaches have been favoured: firstly, analysing released glycans by mass spectrometry; and secondly, analysing whole glycopeptides by mass spectrometry.

Behrens *et al.* (2016) used a quantitative approach to glycopeptide analysis. They first used a HILIC-UPLC-MS approach with fluorescently labelled glycans to create a database of the glycans present on the BG505 SOSIP.664 trimer. This was then used as a benchmark for a subsequent mass spectrometry analysis in order to quantify the levels of glycoforms at each site. Panico *et al.*, (2016) used both a glycan release and a nonquantitative glycopeptide MS approach on a virion-derived gp120 monomer, listing the glycoforms present at each site. Importantly, while mass spectrometry can provide information on the quantities of each type of glycan residue present (mannose, galactose, N-acetylglucosamine, sialic acid or fucose), it loses predictive power as the degree of glycan branching increases, since the specific branches and rotational isomers (or rotamers) allocated are based on assumptions about glycan processing. (Leymarie and Zaia, 2012).

Cao *et al.* (2017 & 2018) used a glycan release approach, which rather than identifying each individual glycoform, instead identified the broad classes of glycans present at each site. Using PNGase F and EndoH, they could differentiate between complex and high mannose or hybrid glycan types, as well as distinguishing between glycosylated and non-glycosylated PNGSs. This allowed sensitive detection and complete coverage. Cao *et al.* (2017) focused on recombinant SOSIP forms from various clades, while Cao *et al.* (2018) compared SOSIP, membrane-bound, pseudovirus-derived and virion-derived Env trimers from different strains. Cao *et al.* (2018) found that viral-derived Env results in glycan shields that are composed of a larger percentage of complex glycans than is seen on the SOSIP trimers; this may be a challenge to be overcome in immunogen design.

In spite of the different approaches and Env sources, these studies were largely in agreement, and remained consistent with previous research that had identified high mannose areas of gp120, such as the intrinsic mannose patch around glycan 332; and complex glycans on gp41, such as at sites 611 and 618 (Doores *et al.*, 2010; Pritchard, Vasiljevic, *et al.*, 2015). These studies also highlight the heterogeneity present on the HIV-1 Env glycan shield. However, in spite of the amazing progress that has been made by these studies, the immense variation of the HIV-1 Env means that only a small sample of the population of millions of varieties of Env has been sampled by these studies.

1.5.4. Molecular Dynamics

Since glycans are difficult to study experimentally due to their heterogeneity (See Section 1.4), they are commonly studied through computational means in order to complement experimental results. Molecular dynamics (MD) is a computational method that uses Newton's laws of motion to apply force fields to theoretical models in order to create representative simulations of a molecule's movement on a nanosecond timescale (Durrant and McCammon, 2011).

MD has been successfully used in HIV-1 Env studies to broaden our understanding of the dynamics of the HIV-1 Env glycoprotein. Stewart-Jones *et al.* (2016) carried out MD studies in conjunction with attempting to crystallise fully-glycosylated Env structures. This was useful in beginning to understand the dynamic interactions of glycans, and highlighting that most known bNAbs need to accommodate glycans within their epitopes. Additionally, the importance of hydrogen bonding to the glycan shield was highlighted. Lemmin *et al.* (2017) carried out a 2 microsecond-long MD simulation on HIV-1 Env. The unprecedented length of the simulation used in this study enabled the characterisation of different structural clusters, which highlighted the asymmetric nature of the HIV-1 Env trimer, as well as its potential for dynamic scissoring and opening movements, which influence access to the CD4 binding site. The phenomenon of scissoring was characterised by the movement of protomers within the same plane, like scissor blades moving together and apart, while the opening phenomenon consisted of each protomer moving outwards from the centre of the trimer. This finding of scissoring differs from what was observed in the 500 ns simulations carried out by Ferreira *et al*. (2018) – though whether this is due to the different structures used, or the shorter length of this study is unclear.

Previous MD studies of HIV-1 Env have typically used uniformly high mannose glycosylated trimers – generally Man5 or Man9, although $Man_7GlcNAc_2$ glycans have also been used (Wood *et al.*, 2013; Gorman *et al.*, 2016; Stewart-Jones *et al.*, 2016; Yokoyama *et al.*, 2016; Lemmin *et al.*, 2017; Li *et al.*, 2017; Moyo *et al.*, 2017; Zhou *et al.*, 2017; Ferreira *et al.*, 2018). While this lines up with the results available from crystallisation studies using kifunensine and GnTI-deficient cell lines, it is not fully representative of the heterogeneity of the HIV-1 glycan shield. Now with more mass spectrometry results available, we can begin to carry out molecular dynamics studies using a mixed set of glycans that better represent the glycan shield.

This project aimed to add to the body of work on the HIV-1 glycan shield, by running molecular dynamics on an HIV-1 Env trimer glycosylated with a mix of glycans representative of the processing levels seen in mass spectrometry. Using recent mass spectrometry papers to inform glycan specificities at each PNGS (Behrens *et al.*, 2016; Panico *et al.*, 2016; Cao *et al.*, 2017), a mixed-glycan model was created for molecular dynamics, which was compared to a model uniformly glycosylated with Man9. Simulations of 500 ns were generated for both systems, and comparative analysis was carried out.

The primary aim of this project was to investigate how different types of glycans influence interactions with the protein and surrounding glycans. This involved the comparison of the dynamics of a Man9 glycosylated trimer with those of a heterogeneous mixed-glycan glycosylated trimer.

The secondary aim of this project was to estimate the potential for glycans to be further processed from

the Man9 starting glycan. This was done by carrying out enzyme and solvent accessibility analysis, and comparing accessible sites with those which are known to be processed in published works.

2. Materials and Methods

Failure is simply the opportunity to begin again, this time more intelligently ~ Henry Ford

2.1. Model Building

The 4TVP crystal structure of an HIV-1 BG505.SOSIP protomer (Pancera *et al.*, 2014) was chosen as the template sequence to model the HIV-1 trimer (Supplementary Figure 1). This choice was made because BG505.SOSIP is the leading Env immunogen, and because it allowed comparison with the results presented in Cao *et al*. (2017). Since the available structures of HIV-1 Env have missing structural information, such as in the gp41 region, multiple PDB files were used so that the structural information could be supplemented from additional files when it was not present in 4TVP; for example inclusion of the 5FUU structure (Lee, Ozorowski and Ward, 2016) was important for the gp41 structural information, which was missing in other files. Additionally, it was important to include the 4NCO trimer (Julien, Cupo, *et al.*, 2013) in order to model a trimer rather than a monomer. The trimerised constructs of 4TVP and 2B4C were also included, though these were inferred, rather than directly taken from a trimeric crystal structure as with 4NCO. Thus, PDB IDs 2B4C (Huang *et al.*, 2005), 4TVP (Pancera *et al.*, 2014), 4NCO (Julien, Cupo, *et al.*, 2013) and 5FUU (Lee, Ozorowski and Ward, 2016) were used as structural templates in order to produce a complete model of the HIV-1 Env ectodomain. (Structures available from: http://www.rcsb.org/.) While the fidelity of these structures is reliant on the quality of crystallisation experiments, using multiple templates allows a degree of averaging across extreme (and biologically questionable) variations; furthermore, Env is a dynamic molecule and the structures formed by the modelling software present atomically (and likely biologically) viable alternative conformations of the glycoprotein.

After aligning the 4TVP template sequence (Pancera *et al.*, 2014), and the 2B4C (Huang *et al.*, 2005), 4TVP (Pancera *et al.*, 2014), 4NCO (Julien, Cupo, *et al.*, 2013) and 5FUU (Lee, Ozorowski and Ward, 2016) sequences using SeaView (Gouy, Guindon and Gascuel, 2010), the theoretical structure of the HIV-1 trimer was computationally modelled using Modeller 9.18 (Šali and Blundell, 1993). Modeller 9.18 works by creating statistical models that fulfil the restraints imposed by the alignment of provided structures, and, thus, for each modelling run the generated models differ slightly in their conformations. In order to assess multiple conformations of the theoretical model for optimal glycosylation, Modeller was run ten times to produce ten models.

2.2. Computational Glycosylation

The ten models were duplicated, and one set of ten uniformly glycosylated with Man9 glycans at every PNGS[2] (Table 1, Figure 5A). The second set of ten was glycosylated with a mix of Man5 glycans, Man9 glycans, and bi-antennary sialylated complex glycans (Complex-glycans), to form mixed-glycan trimers[3] (Table 1, Figure 5B). The Glycam condensed notation for each glycan used can be found in Table 1.

Table 1: Glycans used for computational glycosylation, in abbreviated and expanded forms

Abbreviation	**Glycam Condensed Notation[4]**
Man9	DManpa1-2DManpa1-6[DManpa1-2DManpa1-3]DManpa1-6[DManpa1-2DManpa1-2DManpa1-3]DManpb1-4DGlcpNAcb1-4DGlcpNAcb1-OH
Man5	DManpa1-6[DManpa1-3]DManpa1-6[DManpa1-3]DManpb1-4DGlcpNAcb1-4DGlcpNAcb1-OH
Complex-glycan	DNeup5Aca2-6DGalpb1-4DGlcpNAcb1-2DManpa1-6[DNeup5Aca2-6DGalpb1-4DGlcpNAcb1-2DManpa1-3]DManpb1-4DGlcpNAcb1-4DGlcpNAcb1-OME

For the purposes of characterising the glycans in this thesis, each terminal end (apart from the end attached to the protein) describes a branch: thus, the Man9 and Man5 glycans have three branches, and the Complex-glycan has two branches (Figure 5B). The glycans are further characterised by their length, taken at the longest point, and counted using the distinct levels of sugar residues in the schematics (Figure 5B). Thus, the Man9 glycan is six sugar residues long, the Man5 glycan is five sugar residues long, and the Complex-glycan is seven sugars long. These characteristics are used in combination with the schematic depictions in order to better describe and understand the influence of the different glycans on the dynamics of Env.

Glycan structures were downloaded from Glycam-Web (Woods Group, 2005), and a prototype tool under development for the Glycam-Web suite of tools was used to computationally attach each glycan to particular amino acids of the protein structures generated by Modeller 9.18 (Šali and Blundell, 1993); this was done by exploring different torsional angles (Petrescu *et al.*, 2004) and rotating the glycan to avoid clashes with other parts of the protein or adjacent glycans (software provided by Dr

[2] These models are henceforth referred to as "Man9-Trimers".
[3] Henceforth referred to as "Mixed-Trimers".
[4] See http://glycam.org/docs/custombuilders/condensed-notation/

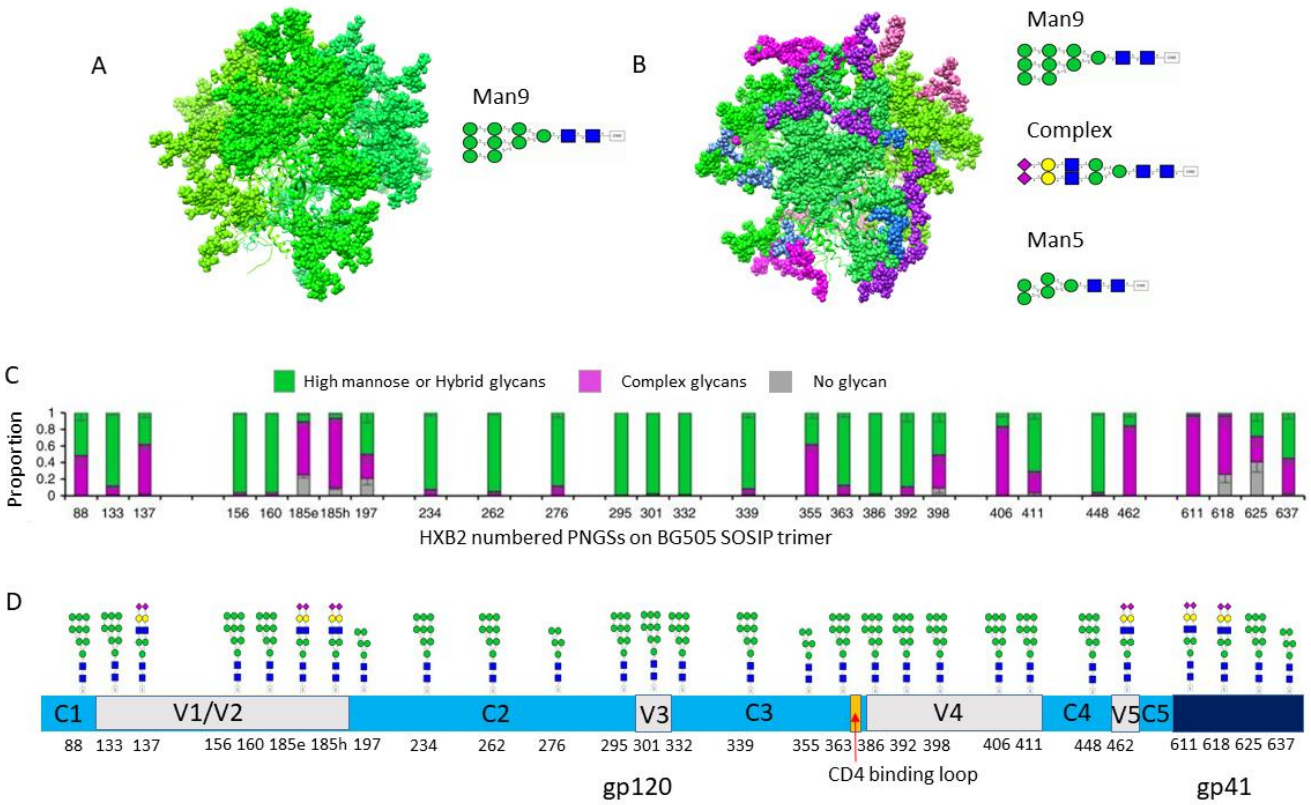


Figure 5: Differential glycosylation of trimers

A and B: Glycosylated trimers of (A) Man9-Trimer, and (B) Mixed-Trimer. Man9 glycans (green), Man5 glycans (blue), and Complex-glycans (purple) and related CFG notation of each glycan are shown, and monomers are differentiated by shades of green. C: Proportion of glycan types found using mass spectrometry for every PNGS on the BG505 SOSIP trimer, adapted from Cao *et al.* (2017). The x-axis depicts PNGS positions labelled according to the reference sequence, HXB2, and the y-axis shows the proportion of the total glycans. D: The glycoform at each PNGS of the Mixed-Trimer, along with HIV-1 Env constant (C1-C5), variable (V1-V5) and gp41 ectodomain regions (not to scale).

Oliver Grant; personal communication). A caveat of this method is that it does not consider every possible angle and position and thus future advancements may improve the computational glycosylation process. However, the glycan shield is so variable that the computational trimers that were generated still reflect what has been found biologically through mass spectrometry experiments (Cao *et al.*, 2017, 2018; Yu *et al.*, 2018). Thus, even though glycans were not added to every PNGS, there is no reason to assume that each PNGS of a particular Env glycoprotein is always glycosylated, which is an important, and often overlooked, point to remember in all glycosylation studies.

The glycosylation patterns of HIV-1 Env trimers are extremely heterogeneous (See Section 1.4) (Pritchard, Harvey, *et al.*, 2015) and thus the representation chosen for the Mixed-Trimer in this study was based on a broad classification of glycans in an attempt to limit this heterogeneity. A small mix of well described glycans that are known to be present on different isolates was used.

Those sites that have been shown by MS studies to be mainly occupied by high mannose glycans were represented using a Man9 glycan. Sites that had a mix of high mannose, hybrid and complex glycans were represented by Man5, to indicate the intermediate levels of processing. Sites that were found to be occupied by complex glycans were represented by the Complex-glycan (Table 1). The type of complex glycan that was chosen represents a large, branched glycan that has been described in previous mass spectrometry results with both a fucosylated and unfucosylated core (Behrens *et al.*, 2016; Stewart-Jones *et al.*, 2016).

The specific glycans added to each PNGS of the Mixed-Trimer (Figure 5D, Table 2) were defined according to the information found in three recent papers (Behrens *et al.*, 2016; Panico *et al.*, 2016; Cao *et al.*, 2017). The information from Cao *et al.* (2017) was used to make a broad differentiation between high mannose and complex glycans (Figure 5C), while the data from Behrens *et al.* (2016) and Panico *et al.* (2016) was used to provide consensus on variable sites, such as site 355, where an almost equal proportion of oligomannose/hybrid and complex glycans was observed (Figure 5C) (Cao *et al.*, 2017); and to distinguish the level of processing, for example at site 197, which was assigned a Man5 glycan given the information in Behrens *et al.* (2016) and Panico *et al.* (2016). Where there was not enough information, or conflicting information, such as at site 406, a Man9 glycan was added to that site as the default.

As mentioned in Section 2.1, the ten models produced by Modeller 9.18 (Šali and Blundell, 1993) each differed slightly in their protein conformations, and accordingly, the number of glycans, out of the 84 available PNGSs, added to each model varied due to different clashes and spatial availability. This is similar to what we would expect *in vivo*, where variable glycosylation occurs (Yu *et al.*, 2018). Since the focus of this study was to compare the glycan-glycan interactions, the trimer model with the highest number of successfully glycosylated PNGSs was taken forward to molecular dynamics. As the underlying protein structure of the MD models was the same, the only differences were in their glycan shields; 80 glycans were successfully added to the final Man9-Trimer, and 79 to the final Mixed-Trimer.

Table 2: Glycoforms at each PNGS on the Man9-Trimer and Mixed-Trimer

HXB2 Sequence Number	Glycoform in Man9-Trimer	Glycoform in Mixed-Trimer
88	Man9	Man9
133	Man9	Man9
137	Man9	Complex
156	Man9	Man9
160	Man9	Man9
185e	Man9	Complex
185h	Man9	Complex
197	Man9	Man5
234	Man9	Man9
262	Man9	Man9
276	Man9	Man5
295	Man9	Man9
301	Man9	Man9
332	Man9	Man9
339	Man9	Man9
355	Man9	Man5
363	Man9	Man9
386	Man9	Man9
392	Man9	Man9
398	Man9	Man9
406	Man9	Man9
411	Man9	Man9
448	Man9	Man9
462	Man9	Complex
611	Man9	Complex
618	Man9	Complex
625	Man9	Man9
637	Man9	Man5

2.3. System Preparation

The two final glycosylated trimers were set up in separate MD systems using the AMBER 14 simulation package (Case *et al.*, 2014). The ff14SB (Maier *et al.*, 2015) and GLYCAM_06-j1 (Kirschner *et al.*, 2008) force fields were used to describe the forces of attraction and repulsion for protein and glycan atoms, respectively. The glycoproteins were computationally immersed in octahedral water boxes with a 15 Å buffer using the TIP5P explicit water model (Mahoney and Jorgensen, 2000), and Na^+ and Cl^- ions were added to neutralise the systems. Prior to solvation, the system sizes were 48 618 atoms for the Mixed-Trimer and 49 197 atoms for the Man9-Trimer. The final system sizes were 660 755 atoms for the Mixed-Trimer and 635 661 atoms for the Man9-Trimer.

2.4. Molecular Dynamics Simulations

The AMBER 14 Molecular Dynamics package (Case *et al.*, 2014) was used for simulations. With the glycoprotein restrained at alpha-carbon atoms by a 5.0 kcal/mol force, energy minimisation was carried out using the steepest descent method for 10 000 cycles, followed by the conjugate gradient method for another 10 000 cycles. Restraints were maintained for the first equilibration step where pressure was kept constant and, using Langevin dynamics, the system was heated to 300 K (NPT). A second equilibration was performed, keeping both temperature and pressure constant (NVT), and releasing the positional restraints in order to relieve bad contacts. Finally, simulations of 500 ns were produced for each system, using a 2 femtosecond (2×10^{-15} second) time step. At the onset of this project, 500 ns was the longest simulation in publication (Stewart-Jones *et al.*, 2016; Zhou *et al.*, 2017), and served as a useful standard for a suitable simulation length to aim for in the present study.

2.5. Analysis

The CPPTRAJ analysis program (Roe and Cheatham, 2013) from the AMBER package (Case *et al.*, 2014) was used to calculate the root mean squared deviation (RMSD), root mean squared fluctuation (RMSF), and the number of hydrogen bonds for each simulation.

Further hydrogen bond analysis was carried out using a combination of Python3 scripts, particularly the functions available in the Pandas module (McKinney, 2010), and Linux command line tools. The number of bonds between pairs of glycans was calculated (Supplementary Methods A); from this, hydrogen bonds and glycans that are important to maintaining the glycan shield, or differed substantially between the trimers (Supplementary Methods B), as well as their corresponding regions on the trimer (Supplementary Methods C) could be identified. Hydrogen bonding longevity was analysed by the percentage presence or absence of a bond over the simulation (Supplementary Methods D), and glycans that were present in hydrogen bonds for the greatest time were identified (Supplementary Methods E). Finally, by filtering the results from the previous hydrogen bond methods, interprotomer interactions, and differences in hydrogen bonding according to glycoform could be investigated.

CPPTRAJ (Roe and Cheatham, 2013) was used to take snapshots of the simulations at 5 ns intervals, which were used for solvent, antibody and enzyme accessible surface area (SASA, AbASA, EASA) analyses with Naccess (Hubbard and Thornton, 1993), each described in more detail below.

In order to investigate the impact of the different glycoforms on accessible surface area, the difference between the accessible surface areas of a glycosylated trimer (Man9-Trimer or Mixed-Trimer) and the same trimer with its glycans removed, was calculated for each snapshot. A probe of radius 10 Å, as

approximation of an antibody loop (Lemmin *et al.*, 2017), was used to calculate the AbASA, and for the SASA, a probe of radius 1.4 Å (commonly accepted as the radius of water) was used. The ten protein residues with the largest median change in accessible surface area between the glycosylated and de-glycosylated forms were plotted for each trimer.

2.6. Investigating EASA and SASA as Proxies for Glycan Processing

Since determining the level of glycan processing from MS results is a costly and time-consuming process which relies heavily on biosynthetic assumptions, the secondary aim of this project was to explore a computational method for determining whether a Man9 glycan will be processed. In its biological context, Man9 is processed by ER Man I, and it is known that glycan-dense regions where enzymes cannot fit are left underprocessed (See Section 1.4).

Thus, the accessible surface area to ER Man I was investigated using Naccess (Hubbard and Thornton, 1993). In order to determine the probe size, the mechanism of ER Man I was taken into account; as a barrel shaped enzyme, only one side needs to fit between glycans for processing to occur. Accordingly, one side of ER Man I (PDB 1X9D) (Karaveg *et al.*, 2005) was measured in UCSF Chimera (Pettersen *et al.*, 2004) and from this it was determined that a probe of 13 Å should be used for the Naccess calculations. These calculations were carried out on the Man9-Trimer, since the Man9 glycan accessibility both in the ER and after oligomerisation and translocation to the Golgi apparatus is believed to influence glycan processing (Doores *et al.*, 2010). Since ER Man I removes one mannose residue at a time, the EASA value of the terminal (0MA) glycan residues were filtered from the Naccess results and binned according to the corresponding processed glycoform of each site on the Mixed-Trimer – i.e. Man9, Man5, or Complex-glycan (See Supplementary Methods F). The distributions were plotted as boxplots using Matplotlib (Hunter, 2007), and statistically analysed for differences using the Mann-Whitney U test.

In addition to exploring the use of EASA in this context, the commonly used SASA metric was investigated, since this has been linked to the level of glycan processing (Suga, Nagae and Yamaguchi, 2018), and thus would be useful to compare. Using a probe of radius 1.4 Å, the SASA of the terminal (0MA) glycan residues was calculated, binned, plotted, and analysed for differences as above.

2.7. Graphing and Visualisation

UCSF Chimera (Pettersen *et al.*, 2004) was used for molecular visualisation. Flourish (Kiln Enterprises Ltd, 2016), xmgrace (Turner, 1991; Stambulchik, 1998), Matplotlib (Hunter, 2007), R (Ihaka and Gentleman, 1996), and Microsoft Excel were used to make the various graphs.

2.8. Statistical Analyses

Statistical analyses were carried out in Python3 using the SciPy module (Jones *et al.*, 2001). The Mann-Whitney U test was chosen due to the skewed distributions of the data, as well as its ability to be used with unequal sample sizes.

3. Results

The facts are always friendly. Every bit of evidence one can acquire, in any area, leads one that much closer to what is true ~ Carl Rogers

Molecular dynamics is a commonly used approach to study glycans and glycoproteins since these molecules are difficult to study in the laboratory (See Section 1.5.4). In the case of the heavily glycosylated HIV-1 envelope protein, this method has been used with a set of uniform glycans in multiple studies (See Section 1.5.4); however this does not fully reflect the diversity of the glycan forms seen on the native virion (See Section 1.4).

In order to expand on previous work, two models were simulated, using molecular dynamics, and compared: a uniformly glycosylated Man9-Trimer and a more diversely glycosylated Mixed-Trimer (See Section 2.2). The primary aim was to investigate the impact of the different glycoforms on the dynamics of the protein. Sections 3.1 to 3.5 of the Results address this question. The secondary aim was to explore whether a computational approach to predicting the level of glycan processing could be proposed. This is addressed in Section 3.6 of this chapter.

3.1. Glycan Occupancy: Assessing the impact of glycoforms on glycan occupancy

Ten HIV-1 Env trimer models were computationally generated using Modeller 9.18 (Šali and Blundell, 1993) in order to find the optimal conformation for maximal glycosylation.

In order to determine the acceptability of the models, they were assessed by their Discrete Optimised Protein Energy (DOPE) scores, where a lower score indicates a more native-like conformation (Shen and Sali, 2006); and by their GA341 scores, which indicate the match between the sequences within the range of 0 (no match) and 1 (complete match) (Melo, Sánchez and Sali, 2002; John and Sali, 2003). All models had DOPE scores of -204 000 or lower, and GA341 scores of 1.

The ten models were duplicated, and one set of ten was glycosylated with Man9 glycans only, while the other set of ten was glycosylated with a mix of Man9, Man5, and Complex-glycans (See Section 2.2).

There are 84 PNGSs distributed unevenly across the HIV-1 trimer sequence used in this work, with the highest concentrations in the V1/V2 and V4 regions (Figure 5D, Supplementary Figure 1). While the protein structure does affect glycan occupation, since the ten models used for glycosylation were

the same for both the Man9-Trimers and the Mixed-Trimers, differences in the percentage glycan occupation can be attributed specifically to the differences in the types of glycans.

The Man9-Trimer was glycosylated uniformly with Man9 glycans; these have three branches (tri-antennary), and are six sugars long. As well as Man9 glycans, the Mixed-Trimer also contained Man5 glycans, which are tri-antennary, five sugar long glycans, and Complex-glycans, which are bi-antennary, seven sugar long glycans (See Section 2.2). Thus the volume occupied by each of these glycoforms differs according to their composition; this in turn affects the possibility for glycan occupancy at any given PNGS.

Major differences in glycan occupation between the Man9-Trimer and the Mixed-Trimer were observed at two PNGSs: 185h, and 392 (Figure 6A and B). Unsurprisingly, since the glycan at site 185h is a Complex-glycan on the Mixed-Trimer, it was added less frequently than on the Man9-Trimer. Conversely, the glycan at site 392 is added less frequently on the Man9-Trimer; while a Man9 glycan is present at this site on both trimers, the surrounding glycans on the Mixed-Trimer are Man5 glycans, allowing extra space, and thus higher occupancy, on this model compared to the Man9-Trimer.

Furthermore, site 185h falls into the V1/V2 region of the HIV-1 Env trimer, while site 392 falls into the V4 region, close to the CD4-binding site. Both regions are densely glycosylated, which may result in greater competition for space in these regions.

After duplicating the models and glycosylating each set with either Man9 glycans or a mix of glycans, the Man9-Trimer and Mixed-Trimer with the most glycans were taken forward for the rest of the

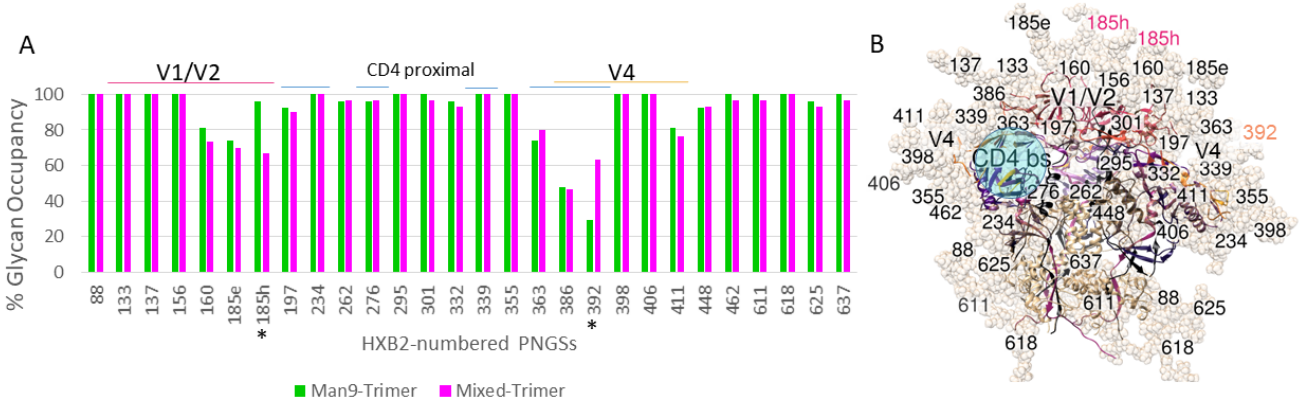


Figure 6: Percentage glycan occupation at each PNGS across ten models

A: Percentage glycan occupation of each PNGS. Two PNGSs of interest, at sites 185h and 392, are starred (*) to indicate where the glycan occupation differs between the Man9-Trimer and the Mixed-Trimer. B: Glycosylated trimer illustrating the clustering of glycans 185h (pink) and 392 (orange), in 3D, rather than linear, space. The CD4-binding site is show in blue for structural context.

analysis. Out of 84 PNGSs, the Man9-Trimer had 80 N-linked glycans and the Mixed-Trimer had 79 N-linked glycans. The absent glycans were distributed across the different protomers (Table 3).

Table 3: Potential N-glycosylation sites that were not glycosylated on the final models

Man9-Trimer	Mixed-Trimer
392 (Protomer 1)	386 (Protomer 1)
363 (Protomer 2)	462 (Protomer 1)
392 (Protomer 2)	185h (Protomer 2)
363 (Protomer 3)	392 (Protomer 2)
	363 (Protomer 3)

3.2. RMSD: Assessing the stability of simulations

The root mean squared deviation (RMSD) of the protein backbone atoms was calculated relative to the first frame for each of the trimers using CPPTRAJ (Roe and Cheatham, 2013) to assess the stability of the simulations (Figure 7). The RMSD represents the overall movement of the molecule at every time point in the simulation.

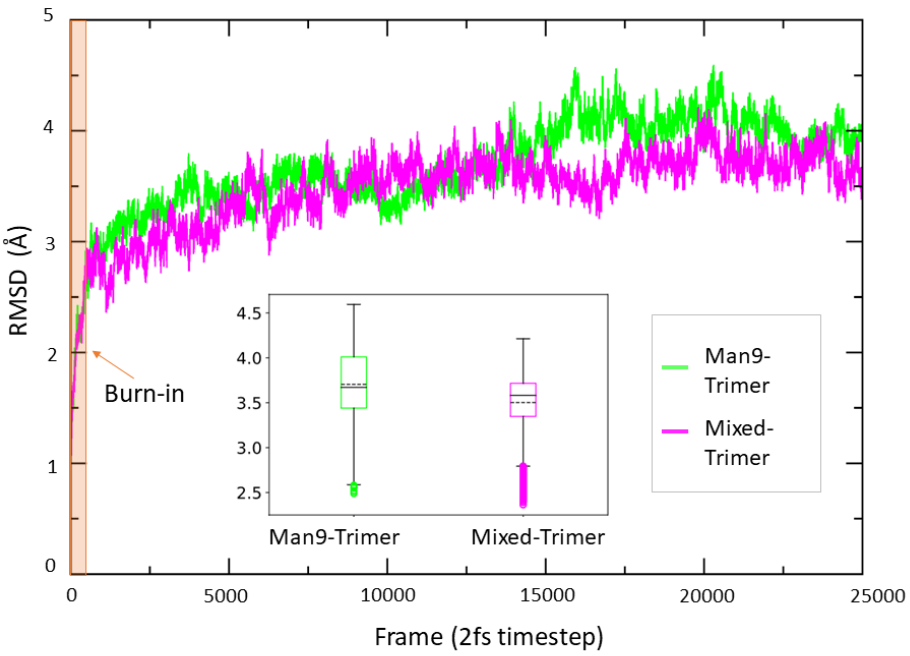


Figure 7: Graph to compare RMSD of the Mixed-Trimer and Man9-Trimer

RMSD (Angstrom) of the Man9-Trimer (green) and Mixed-Trimer (purple). The steep upward gradient represents the burn-in time (shaded orange). Each frame represents 2 femtoseconds (fs; 2 x 10^{-15} seconds) of simulation out of a total of 500 ns of simulation. The included subplot shows boxplots depicting minimum, lower quartile, mean (dotted line), median, upper quartile, and maximum of the RMSD for the Man9-Trimer and Mixed-Trimer (burn-in removed). Points depict outliers.

The steep ascent at the beginning of the graph represents the "burn-in" time, where the molecules are still equilibrating, and a burn-in time of 10ns was removed prior to further analysis of the simulations (Figure 7, shaded orange). The median RMSD values for the Man9-Trimer and the Mixed-Trimer were 3.67 Å and 3.58 Å, respectively. For a large proportion of the simulation (64%), the RMSD of the Mixed-Trimer backbone was lower, and less fluctuating, than that of the Man9-Trimer, indicating a slightly more stable protein form within the Mixed-Trimer simulation.

3.3. RMSF: Investigating the effect of different glycans on protein stability

The root mean squared fluctuation (RMSF) of the alpha-carbon atoms of each protein residue in the trimer was calculated using CPPTRAJ (Roe and Cheatham, 2013) and is a measure of how much each residue moves relative to the average structure over the course of the simulation (Figure 8).

There are notably higher fluctuations towards the beginning of the gp41 protein chains in the first and third protomers, which may represent a confounding factor that should be improved upon by adding greater restraints on these residues in future studies. The outlying RMSF values caused by this issue

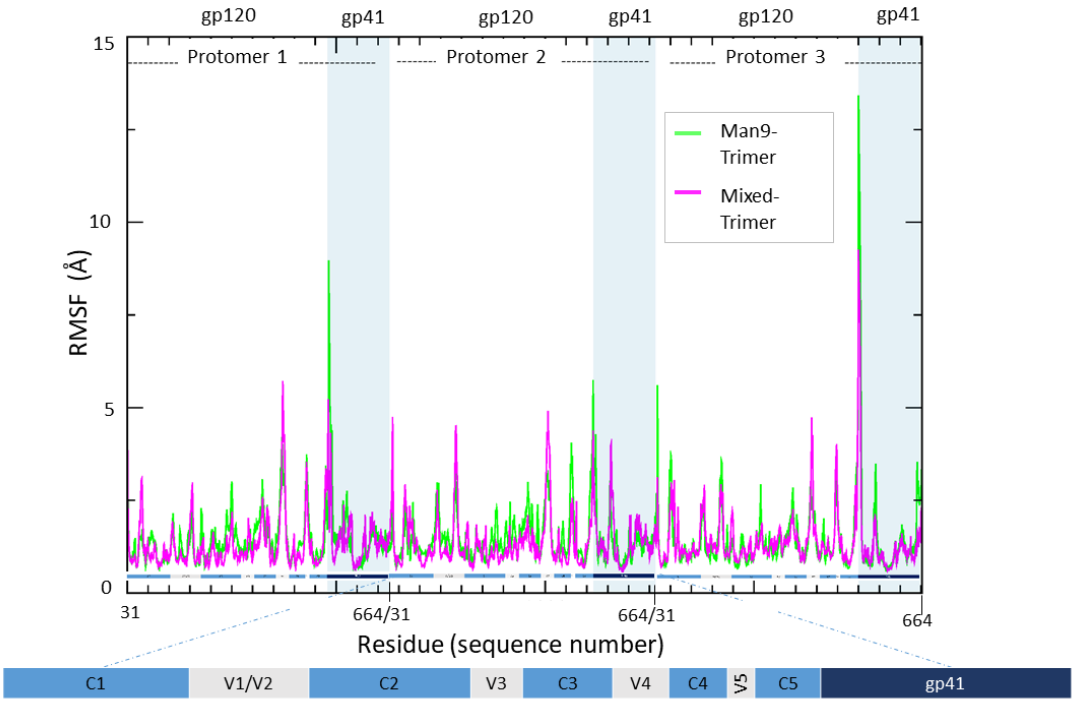


Figure 8: Comparison of RMSF for Mixed-Trimer and Man9-Trimer

RMSF was calculated per protein residue for each trimer to investigate whether the different glycoforms affected protein stability. Protomer divisions, as well as gp120 and gp41 distinctions are indicated above the graph. HIV-1 regions are shown by alternating blocks of blue (constant regions) and light grey (variable regions) for gp120, and dark blue for gp41. This is enlarged at the bottom of the figure for readability.

were excluded for statistical analysis, and only RMSF values greater than or equal to 6.7 were included (Supplementary Table 1).

The median RMSF values for the Man9-Trimer and Mixed-Trimer were 1.25 Å and 1.15 Å respectively. The Man9-Trimer exhibited higher RMSF 55% of the time, indicating a more fluctuating protein backbone in the Man9-Trimer than in the Mixed-Trimer. This corresponds to the findings from the RMSD analysis.

Similar patterns of stability were present across the protomers. Local maxima corresponded to variable loops, with the V4 and V5 regions having the most pronounced fluctuations. Thus, variable regions had a greater range of movement than the constant regions – with the exception of the start of the C1 region. The high fluctuations in the first residues of the C1 regions may be due to the lack of sufficient restraint at the ends of the gp120 subunits.

3.4. Hydrogen Bonds: Investigating interactions in the glycan shield

3.4.1. Quantifying the glycan interactions from simulations

In order investigate the differences in the glycan shields themselves (rather than just the underlying proteins), hydrogen bond analysis was carried out on the simulations using CPPTRAJ (Roe and Cheatham, 2013).

A hydrogen bond is defined as an interaction between a hydrogen bond donor (X-H) – that is, a hydrogen bound to an electronegative atom (such as oxygen, fluorine, or nitrogen) – and a hydrogen bond acceptor (X:), where an atom contains a lone pair of electrons.

In order to quantify the glycan interactions, the number of hydrogen bonds between every interacting glycan pair was calculated, e.g. between glycan 133 of Protomer 1 (1_133_m9) and glycan 137 of Protomer 1 (1_137_m9/c)[5]. Only the interactions with other glycans were included for this count; intra-glycan interactions were specifically excluded (Supplementary Methods A).

[5] This notation will be used to refer to glycans in figures and tables in the rest of the thesis: 1_133_m9 refers to protomer 1, glycan 133, with m9, which is a Man9 glycan. The notation m5 is used to denote Man5 glycans, e.g. 1_197_m5, and c is used to denote Complex-glycans, e.g. 1_137_c. Where the same glycan takes on different glycoforms in the two trimers and is being discussed in a general manner, both glycoforms will be included, e.g. 1_137_m9/c refers to glycan 137 of protomer 1, which is a Man9 glycan in the Man9-Trimer, and a Complex-glycan in the Mixed-Trimer. The specific inclusion of the glycoform whenever a glycan is referred to is to aid the reader in following the respective types of glycans, which change between the two trimers.

The Mixed-Trimer had a total of 1 332 406 hydrogen bonds over the course of the simulation, 24.8% more than the Man9-Trimer, where 1 067 251 hydrogen bonds occured. By counting the number of lone pairs and hydrogens bound to electronegative atoms, the total number of hydrogen bonds possible over the simulation (if all potential hydrogen bond donors and acceptors were involved at every frame) was coarsely determined in order to obtain estimates for the Mixed-Trimer and Man9-Trimer (Table 4). The maximum for the Man9-Trimer was estimated to be 66 640 000 potential hydrogen bonds, whereas 61 691 000 potential hydrogen bonds were estimated for the Mixed-Trimer. By this calculation, the Man9-Trimer has the potential to form 8% more hydrogen bonds than the Mixed-Trimer. Thus, the observation here, where the Man9-Trimer had substantially less hydrogen bonds present across the simulation than the Mixed-Trimer, is remarkable.

Table 4: Potential for hydrogen bonds

Glycan type	**Hbond acceptors X:**	**Hbond donors X-H**	**Count for Man9-Trimer**	**Count for Mixed-Trimer**
Man9	112	34	80	51
Man5	72	22	0	12
Complex-glycan	128	36	0	16
Total potential Hbonds in 1 frame:*			2720	2518
Total potential Hbonds across simulation (24500 frames):			66 640 000	61 691 000

*The total potential hydrogen bonds were calculated based on the number of Hbond donors, since this is the limiting factor.

On the Man9-Trimer, two sets of hydrogen bonded glycan pairs (glycans 88 and 625 on Protomer 3, and glycans 295 and 332 on Protomer 1) (Figure 9A, Supplementary Figure 2) made up 12.8% of its overall hydrogen bonding. The Mixed-Trimer on the other hand had a more distributed hydrogen bond count across interactions, with 8% more individual bonds making up the total hydrogen bond count. The two glycan pairs with the highest hydrogen bond count on the Mixed Trimer (Man9 glycans 262 and 448 on Protomer 1, and the 234 Man9 glycan and 462 Complex-glycan on Protomer 2) made up 7.4% of the total hydrogen bonding.

Each glycan bonding pair contained two glycans, and a number of glycans were involved in multiple interactions. Due to this finding, bonding pairs were divided into their component glycans in order to determine which glycans formed the largest number of hydrogen bonds over the course of the simulation, e.g. the interaction of 72 361 hydrogen bonds between glycans 295 and 332 of Protomer 1 of the Man9-Trimer would result in 72 361 being added to the total contribution by glycan 295, as well as that of glycan 332 (Supplementary Methods B).

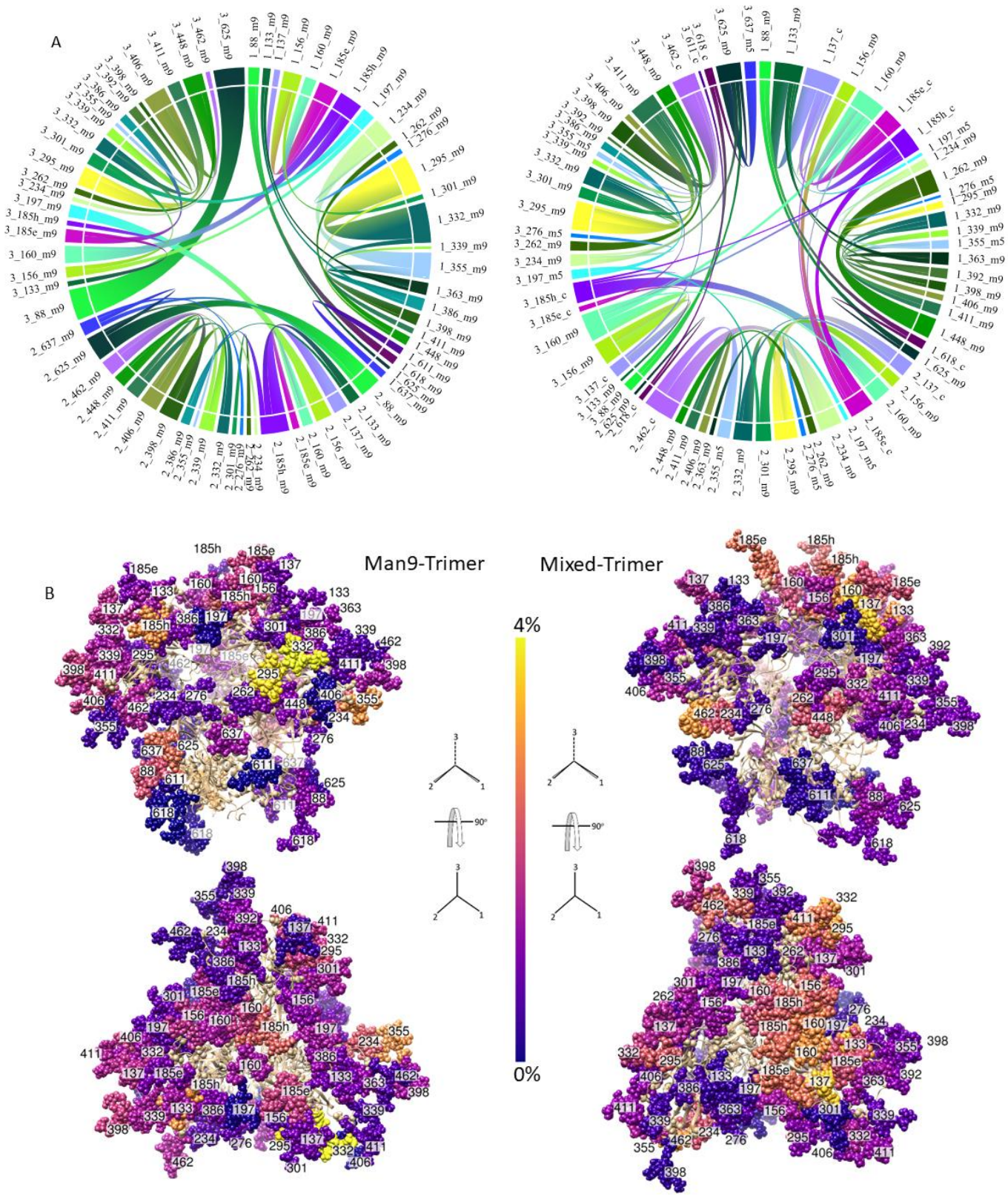


Figure 9: Comparison of inter-glycan hydrogen bonding for the Man9-Trimer and Mixed-Trimer

A: Chord diagrams of glycan-glycan hydrogen bond counts ≥5000 coloured according to glycoform in Mixed-Trimer and repeated across protomers. Green to yellow range = Man9, blue range = Man5, Purple range = Complex-glycans. Each chord joins 2 glycans and its width corresponds to the total number of hydrogen bonds present between those glycans across the simulation. B: Glycan percentage contribution to overall hydrogen bonding. Blue (0%) to yellow (4%). Individual percentages and hydrogen bonding counts by glycan can be found in Supplementary Table 2.

The hydrogen bond contribution for every glycan was converted into a percentile in order to allow comparison between the trimers (Figure 9B, Supplementary Table 2). In this way, the glycans that were most involved in hydrogen bonding could be identified. This allows us to see not only which bonds played an important part in stabilising the glycan shield, but which individual glycans too.

On the Man9-Trimer, it can be seen that the glycans involved in the two dominant hydrogen bonds (glycans 332 and 295 of Protomer 1, and 625 and 88 of Protomer 3) also had the largest contribution to the overall hydrogen bonding in the glycan shield (Figure 9B, Supplementary Table 2).

This was not the case for the Mixed-Trimer, where the glycans that contributed most to the hydrogen bonding were involved in numerous different bonding pairs, e.g. the Complex-glycan at site 137 of Protomer 1 of the Mixed-Trimer had a total of 91 363 hydrogen bonds – including 43 870 bonds with Man9 glycan 133 of Protomer 1, 24 574 with Complex-glycan 185e of Protomer 1, 13 899 with Man9 glycan 156 of Protomer 1, and 7 227 bonds with Complex-glycan 185e of Protomer 2, as well as minor bonds with other glycans (Figure 9A, Supplementary Figure 3). The two highest contributors to hydrogen bonding on the Mixed-Trimer were Complex-glycans.

Symmetry across the protomers did not exist when it came to identifying individual glycans that consistently formed hydrogen bonds with surrounding glycans, e.g. while the Complex-glycan 137 on Protomer 1 of the Mixed-Trimer had a high hydrogen bond count and contributed 3.4% of the total hydrogen bonding, the same glycan on Protomers 2 and 3 only contributed 1.6% and 1.4% respectively. This non-symmetrical sharing of hydrogen bonds between protomers was seen on both the Man9-Trimer and the Mixed-Trimer.

The apex glycans contributed to the hydrogen bonding to a higher degree on the Mixed-Trimer, where they made up about a fifth of the total, compared to the apex glycans on the Man9-Trimer, where they accounted for less than a third of the total hydrogen bonding (Figure 9B). On the Mixed-Trimer, these glycans at the apex were predominantly Complex-glycans.

3.4.2. Assessing how glycan-glycan hydrogen bonds cluster into the defined regions of HIV-1 Env

In order to investigate if hydrogen bonds were more likely to form across specific regions of the HIV-1 trimer, the total number of glycan-glycan hydrogen bonds was determined for the variable (V1-V5) loops and constant (C1-C5) regions of gp120, as well as for gp41 (Supplementary Methods C). Heatmaps were then plotted to show regions with higher hydrogen bond interactions within each trimer (Figure 10A and B).

It is likely that areas with a high density of glycans, or glycans in close proximity, would interact the most. Thus, the major diagonal ascending from left to right (Figure 10A and B), where each protomer of the trimers is interacting within itself, is unsurprising. Instead, it is in the further details that interesting differences between the trimers can be seen.

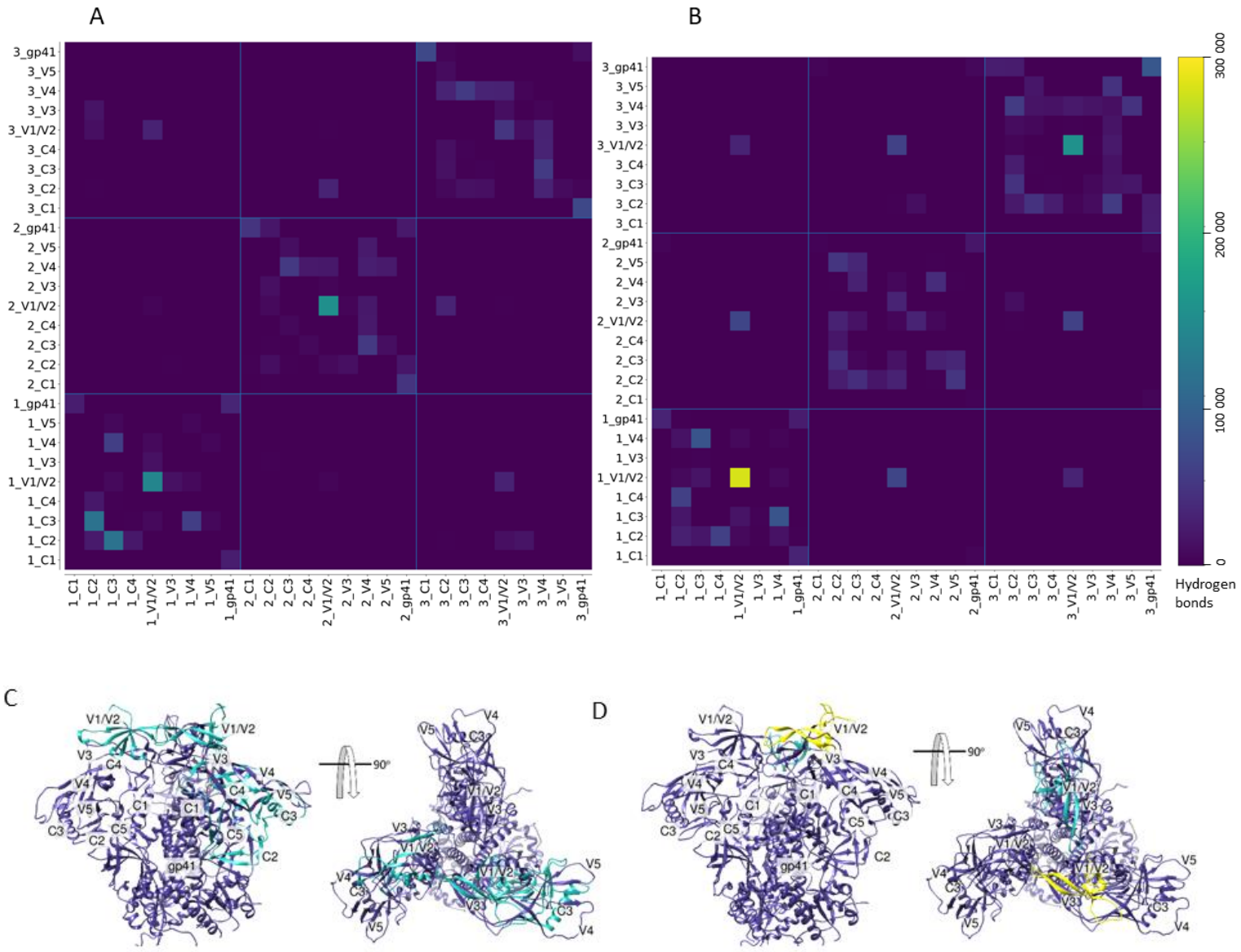


Figure 10: Comparing glycan-glycan hydrogen bonds by HIV-1 region

Individual hydrogen bonds were allocated to their respective HIV-1 regions to investigate which regions were most involved in hydrogen bonding. A and B: Heatmaps of the Man9-Trimer (A) and Mixed-Trimer (B) showing hydrogen bonding between HIV-1 regions. Grid used to represent protomer divisions. Main diagonal shows intra-protomer interactions. Note: The Mixed-Trimer heatmap does not contain the V5 region in Protomer 1 as the only glycan in this region was not added during glycosylation (Table 3). Neither heatmap contains the C5 region as it does not contain any glycans. C and D: 3D depictions of HIV-1 regions of high hydrogen bonding involvement in (C) the Man9-Trimer, and (D) the Mixed-Trimer. Notable regions of high hydrogen bonding are highlighted according to the colours of the highest count for these regions in their respective heatmaps.

The most notable hydrogen bonding occurred in the glycan-dense V1/V2 region of Protomer 1 of the Mixed-Trimer, where there was a total of 140 785 intra-region hydrogen bonds (Figure 10B and D, highlighted in yellow). In addition, the V1/V2 region of Protomer 3 of the Mixed-Trimer included a remarkable number of intra-region hydrogen bonds, totalling 75 748 bonds (Figure 10B and D, highlighted in turquoise).

Similarly, the V1/V2 regions of Protomers 1 and 3 of the Man9-Trimer formed comparable levels of intra-region hydrogen bonding to the V1/V2 region of Protomer 3 of the Mixed-Trimer (Figure 10A and C, highlighted in turquoise). Intriguingly, the C2 and C3 regions of Protomer 1 also interacted with each other to a comparable level (Figure 10A and C, highlighted in turquoise); this was the most notable inter-region hydrogen bonding in either trimer.

Interestingly, for both trimers, one protomer (Protomer 2 for the Mixed-Trimer and Protomer 3 for the Man9-Trimer) formed substantially less hydrogen bonds within its V1/V2 region compared to the other two protomers (Figure 10A and B).

In contrast to the high levels of hydrogen bonding formed by the glycan-dense V1/V2 region, the V4 region, which is also glycan-dense, did not form an exceptional level of hydrogen bonds.

Outside of the main diagonal of the heatmaps, we see a distinct difference in interprotomer hydrogen bonding between the two trimers (Figure 10A and B). The Mixed-Trimer formed interprotomer hydrogen bonds mainly between V1/V2 regions, with each protomer interacting substantially with every other protomer. In contrast, the Man9-Trimer formed interprotomer interactions between the C2 and V1/V2, V1/V2 and V1/V2, and C2 and V3 regions of Protomers 1 and 3, as well as interactions between the V1/V2 and C2 regions of Protomers 2 and 3. Interestingly, the Man9-Trimer exhibited hardly any hydrogen bonds between Protomer 1 and 2, totalling 3 158 interprotomer bonds between all regions of these protomers. While the V1/V2 region was involved in the interprotomer interactions on both trimers, the interactions between the V1/V2 loops were more pronounced on the Mixed-Trimer than on the Man9-Trimer, with 150 246 and 33 575 hydrogen bonds respectively.

3.4.3. Investigating differences in interprotomer hydrogen bond interactions

Due to the marked differences in interprotomer interactions between the Man9-Trimer and the Mixed-Trimer, this subset of hydrogen bonds was further investigated, in order to examine the specific glycans and glycan bonding pairs involved.

The Mixed-Trimer had a total of 177 778 interprotomer hydrogen bonds, which is nearly double the 96 551 interprotomer hydrogen bonds found on the Man9-Trimer. These interprotomer numbers

translate to 13.3% and 9% of the total number of hydrogen bonds in the glycan shields of the Mixed-Trimer and Man9-Trimer respectively. Four dominant interprotomer interactions (with a count of 5000 or more hydrogen bonds) were identified for the Man9-Trimer and ten dominant interprotomer interactions were identified for the Mixed-Trimer (Figure 11A, Table 5).

The range of magnitudes of hydrogen bonding between glycan bonding pairs of the dominant interactions was remarkable in its difference between the trimers: the Man9-Trimer had a narrow range, between 11 844 and 26 006 bonds, while the Mixed-Trimer had a wide range, between 6031 and 39 730 bonds (Table 5).

The dominant interprotomer interactions represented 82.1% and 83.4% of the total interprotomer interactions for the Man9-Trimer and Mixed-Trimer respectively, with minor interactions (those with a count of less than 5000 hydrogen bonds) making up the remaining 17-18%. Of significant interest was the observation that eight of the ten dominant interprotomer interactions on the Mixed-Trimer involved Complex-glycans (Table 5).

Interestingly, the Complex-glycan 185e on Protomer 2 of the Mixed-Trimer appears to fill in for an adjacent missing Complex-glycan at position 185h on this protomer (Table 3; Figure 11B). However, the Complex-glycans at position 185e of the other two protomers, where the glycan at position 185h is present, are not involved in the dominant interprotomer interactions (Figure 11A).

Table 5: Dominant interprotomer interactions for the Man9-Trimer and Mixed-Trimer (≥5000)

Man9-Trimer		**Mixed-Trimer**	
Interaction	**Number of Hydrogen bonds**	**Interaction**	**Number of Hydrogen bonds**
2_156_m9 - 3_197_m9	26 006	2_160_m9 - 3_185h_c	39 730
1_185h_m9 - 3_160_m9	25 757	1_185h_c - 2_185e_c	25 400
1_197_m9 - 3_301_m9	15 646	1_156_m9 - 2_185e_c	22 756
1_197_m9 - 3_156_m9	11 844	1_160_m9 - 3_160_m9	13 844
		2_301_m9 - 3_197_m5	10 696
		1_185h_c - 3_160_m9	9 585
		1_137_c - 2_185e_c	7 227
		1_160_m9 - 2_185e_c	6 834
		1_185h_c - 3_185h_c	6 187
		2_618_c - 3_611_c	6 031
Total	**79 253**	**Total**	**148 290**

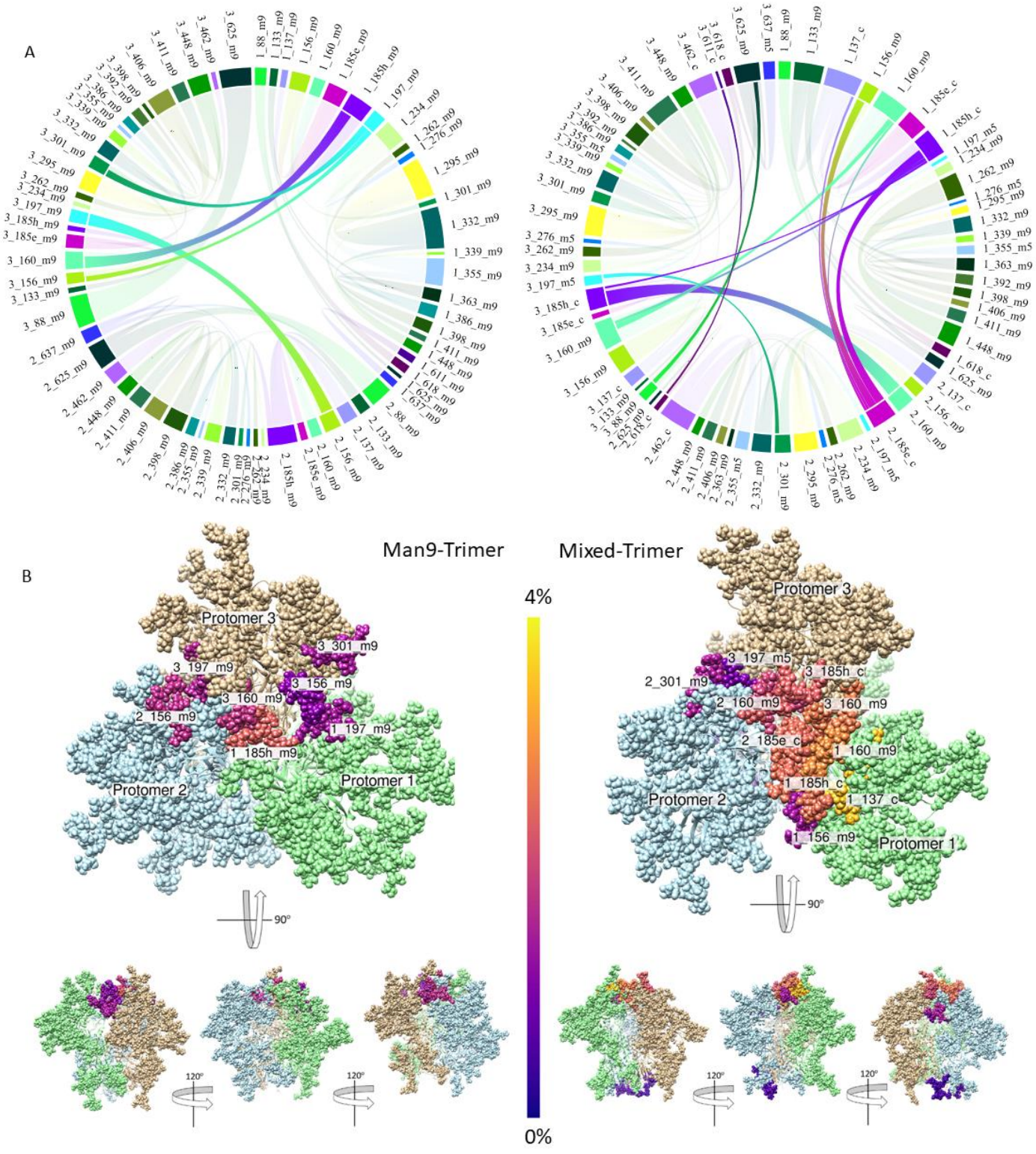


Figure 11: Comparison of interprotomer hydrogen bonding of the Man9-Trimer and Mixed-Trimer

A: Chord diagrams of interprotomer glycan-glycan hydrogen bond counts ≥5000 coloured according to glycoform of the Mixed-Trimer and repeated across protomers. Green to yellow range = Man9, blue range = Man5, Purple range = Complex-glycans. Each chord joins 2 glycans and its width corresponds to the total number of hydrogen bonds present between those glycans across the simulation. B: Top view and 3-way rotated side view of trimers showing glycans that contribute to interprotomer interactions, coloured according to their percentage contribution to overall hydrogen bonding in the glycan shield. Blue (0%) to yellow (4%).

Another striking observation was that for both trimers, the majority of interprotomer interactions were formed by glycans at the apex of the trimer (glycans at sites 137, 156, 160, 185e, 185h, and 301). On the Mixed-Trimer, the glycans present on each individual protomer interacted with the glycans on the other two protomers (Table 5, Figure 11A and B). Conversely, a two-way stabilisation effect was observed in the Man9-Trimer, where the glycans on Protomers 1 and 3, and the glycans on Protomers 2 and 3 interacted (Table 5, Figure 11A and B). This two-way stabilisation is likely linked to the scissoring that was observed in the Man9-Trimer simulation, and which results in an exposed CD4 binding site between Protomers 1 and 2, which do not interact during the simulation (Figure 12).

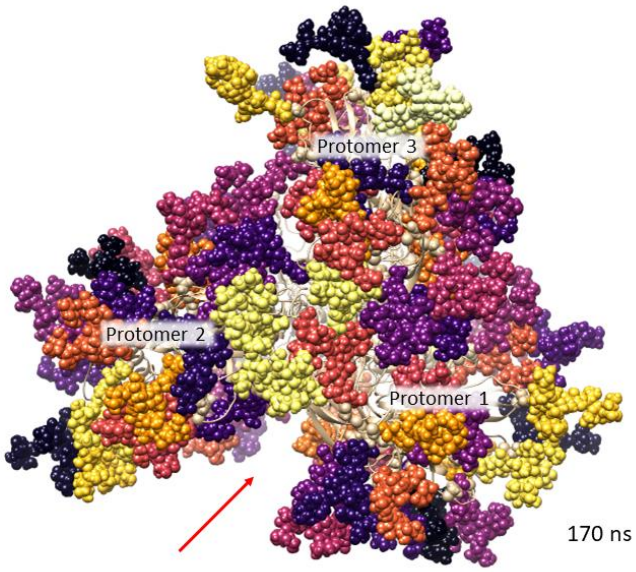


Figure 12: Scissoring of the Man9-Trimer

This snapshot, taken at 170 ns, highlights the scissoring of the protomers on the Man9-Trimer. The red arrow indicates the exposed CD4 binding site between Protomers 1 and 2. Colours here were used to differentiate between neighbouring glycans.

Out of the five glycans involved in interprotomer interactions on both models (glycans at sites 156, 160, 197, 185h, and 301), only glycans 185h and 160 formed the same interacting interprotomer pair on both the Man9-Trimer and the Mixed-Trimer (Table 5). These interactions were observed between Protomers 1 and 3, and for the Mixed-Trimer the interacting pair consisted of a Man9 glycan (site 160) and Complex-glycan (site 185h) (Figure 13). The Mixed-Trimer extended the list of its interprotomer interaction glycans to include the glycans at sites 137, 185e, 611, and 618 – all of which were Complex-glycans (Figure 13).

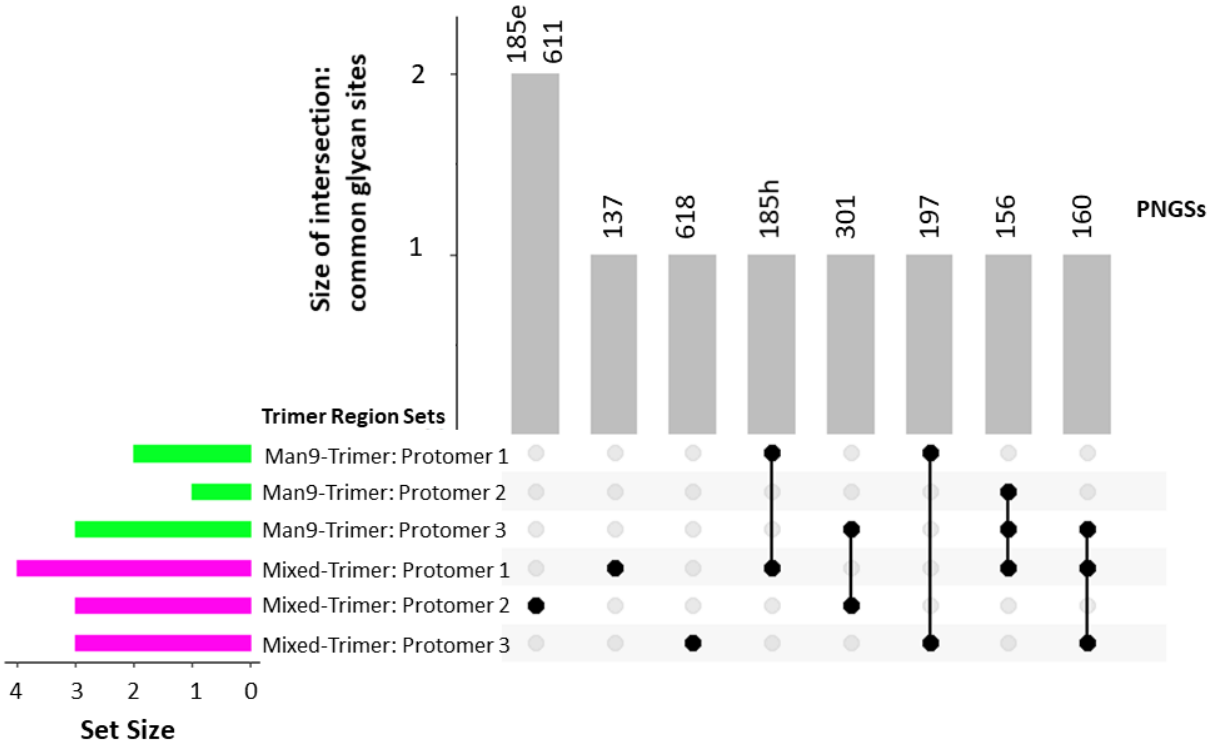


Figure 13: UpSet plot of common glycan sites involved in dominant interprotomer interactions

The dominant interprotomer glycan interactions of the Mixed-Trimer and Man9-Trimer are illustrated; the set size presents the number of glycans from each protomer involved in the dominant interprotomer hydrogen bonding interactions; the size of the intersection shows the number of glycan sites common across protomers and the shaded circles (and lines) indicate the set of protomers where that glycan is involved in interprotomer interactions.

3.4.4. Investigating hydrogen bond longevity versus strength of bonding

Glycans interact dynamically (Stewart-Jones *et al.*, 2016), with some forming lasting interactions over nanoseconds, and others interacting transiently over femtoseconds. Additionally, the number of hydrogen donors and acceptors differs between glycoforms and therefore the potential number of hydrogen bonds varies depending on the glycan pair (Table 4).

Here, hydrogen bonds were analysed according to presence or absence in each 2 femtosecond frame (Supplementary Methods D). Similar to the quantification of hydrogen bonds by glycan (Section 3.4.1), these bonds were then divided into their component glycans in order to identify the glycans that were most involved in hydrogen bonding over the course of the simulation (Supplementary Methods E).

Glycans that were present in hydrogen bonding pairs in more than 90% of the simulation were defined as predominantly present and those present in less than 10% of the simulation were defined as predominantly absent.

The Man9-Trimer had five glycans that were predominantly absent from its set of glycans forming hydrogen bonds (Table 6). The Mixed-Trimer, on the other hand, only had one glycan that was predominantly absent, which was the Complex-glycan 611 on Protomer 1 (Table 6). Notably, the predominantly absent glycans were largely located at the base of the trimer (Figure 14A and B).

Conversely, the Mixed-Trimer had thirteen glycans that were predominantly present in the set of glycans forming hydrogen bonds, while the Man9-Trimer had only seven (Table 6). Strikingly, no Man5 glycans were included in the predominantly present glycans on the Mixed-Trimer. The predominantly present glycans clustered on the Mixed-Trimer (Figure 14B), and conversely were dispersed on the Man9-Trimer (Figure 14A).

Table 6: Glycans that were predominantly present or absent in hydrogen bonding

Predominantly Absent		**Predominantly Present**	
Man9-Trimer	**Mixed-Trimer**	**Man9-Trimer**	**Mixed-Trimer**
1_406_m9	1_611_c	1_185h_m9	1_133_m9
2_197_m9		1_295_m9	1_137_c
2_611_m9		1_332_m9	1_160_m9
2_618_m9		1_355_m9	1_185h_c
3_618_m9		2_625_m9	1_262_m9
		3_88_m9	1_448_m9
		3_625_m9	2_234_m9
			3_160_m9
			3_185h_c
			3_295_m9
			3_332_m9
			3_411_m9
			3_625_c

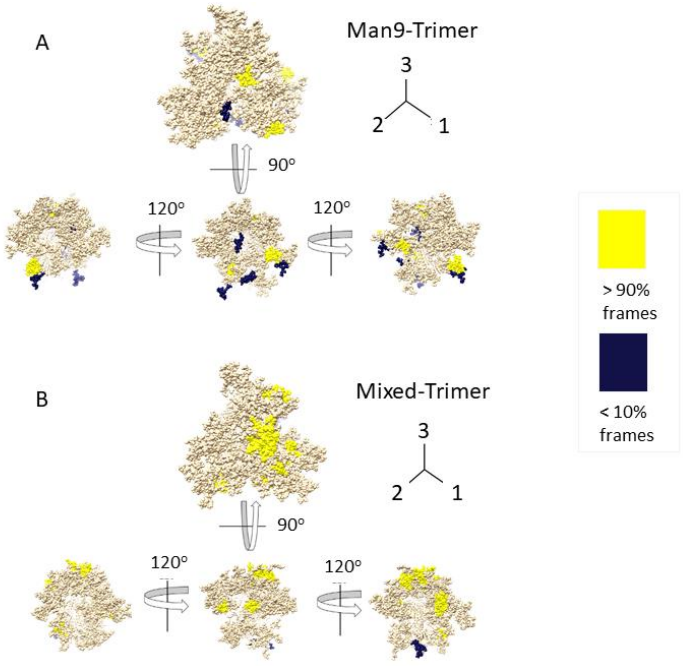


Figure 14: Location of predominantly present and absent glycans in hydrogen bonding

Hydrogen bonds were analysed according to presence or absence by frame, and the glycans present in >90% or <10% of the simulation were highlighted on the trimers. A: Man9-Trimer. B: Mixed-Trimer

3.4.5. Glycoform differences in hydrogen bonding

Since the glycan shields from the two models used in this work are differentiated by their glycoforms, these were compared in the context of hydrogen bonding.

End to end, at the longest point, the Man9 glycan is six sugars long, and it has three branches. Similarly, the Man5 glycan has three branches, but it is five sugars long. The Complex-glycan is seven sugars long and has two branches, and unlike the Man5 and Man9 glycans, it is made up of a mix of sugar residues, including a reactive terminal sialic acid group (See Section 2.2). Due to these differences in structure it was expected that differences in hydrogen bonding would be observed.

The Man5 glycans and Complex-glycans (Figure 4) on the Mixed-Trimer were compared to the Man9 glycans at equivalent positions on the Man9-Trimer (Table 2). Since the Man5 glycans are shorter and have fewer residues than the Man9 glycans, it was anticipated that the glycans at these sites would be less involved in hydrogen bonding on the Mixed-Trimer than on the Man9-Trimer.

The Man5 and Man9 glycans present at sites 197, 276, 355, and 637 were involved in hydrogen bonding in a median of 42.4% of frames on the Man9-Trimer, compared to 34.8% of frames on the Mixed-Trimer. However, the difference in hydrogen bond presence between the Man5 and Man9

glycans at equivalent sites was not significant (Mann Whitney U = 71, $n_{man9} = n_{mixed} = 13$, p = 0.49, one-tailed).

Strikingly, the glycans at position 197 of Protomers 1 and 3 of the Man9-Trimer were involved in three of its four dominant interprotomer interactions. These accounted for 67.5% of the dominant Man9-Trimer interprotomer interactions (Table 7). Conversely, for the Mixed-Trimer, the Man5 glycan at site 197 formed part of only one of the ten dominant interprotomer interactions (site 197 on Protomer 3); this accounted for 7.21% of the dominant Mixed-Trimer interprotomer interactions (Table 7).

Table 7: Comparison of sites where glycoforms differ

Glycoform equivalent sites	**Man9/Man5 equivalent glycans**		**Man9/Complex equivalent glycans**	
Trimer	**Man9-Trimer**	**Mixed-Trimer**	**Man9-Trimer**	**Mixed-Trimer**
Total % glycan contribution to hydrogen bonding	10.90%	7.50%	17.40%	25.20%
Median percentage presence in frames	42.40%	34.80%	44.40%	74.10%
Total hydrogen bond count	217 272	195 622	326 910	596 881
Percentage of interprotomer bonds ≥ 5 000	67.50%	7.21%	32.50%	83.45%

Turning to the Complex-glycans: these were added to sites 137, 185e, 185h, 462, 611, and 618 on the Mixed-Trimer. It was expected that the Complex-glycans would contribute more to hydrogen bonding than the Man9 glycans, due both to their greater length, and the presence of a reactive terminal sialic acid group.

The median percentage of frames where the glycans were involved in hydrogen bonding was 44.4% for the Man9-Trimer, and 74.1% for the Mixed-Trimer (Table 6). This difference was significant (Mann Whitney U = 86, $n_{man9} = 18$, $n_{mixed} = 16$, p = 0.02, one-tailed). Furthermore, the Complex-glycans contributed 7.8% more to the overall hydrogen bonding of the Mixed-Trimer than their Man9 glycan counterparts contributed to the hydrogen bonding of the Man9-Trimer.

The Complex-glycans also played an important role in the interprotomer interactions on the Mixed-Trimer, as was noted in Section 3.4.3. Compared to the glycans at the same sites on the Man9-Trimer,

the Complex-glycans contributed 51% more to the dominant interprotomer hydrogen bonding in the Mixed-Trimer. This was in contrast to the difference observed between the Man5 and Man9 glycans at equivalent sites, mentioned previously.

3.5. Changes in Solvent and Antibody Accessible Surface Area

In order to investigate the impact of different glycans on the accessible surface area of the trimers, the change in solvent accessible surface area (ΔSASA) and antibody accessible surface area (ΔAbASA) between the glycosylated and de-glycosylated forms was calculated for every residue of a Man9-Trimer and a Mixed-Trimer at 5 ns intervals (See Section 2.5).

Naccess (Hubbard and Thornton, 1993) was used for the accessible surface area calculations. A probe of radius 10Å was used to calculate the ΔAbASA (Figure 15A), since 20Å is the width of an antibody loop (Lemmin *et al.*, 2017). The ΔSASA calculations were done with a probe of radius 1.4 Å as this approximates the size of water (Figure 15B) (See Section 2.5).

It is worth noting that the accessible surface area changes for SASA and AbASA are in vastly different ranges due to the differing Gaussian surfaces that are produced in the rolling ball calculation of Naccess (Hubbard and Thornton, 1993) (Figure 15 A and B).

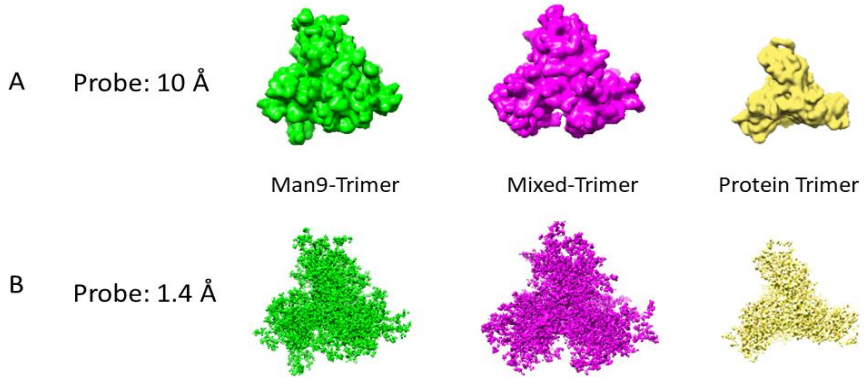


Figure 15: Surfaces for antibody and solvent accessible surface area calculations

A and B: Gaussian surfaces for (A) ΔAbASA (using a 10Å probe) and (B) ΔSASA (using a1.4Å probe) of Man9-Trimer (green), Mixed-Trimer (purple) and the Protein Trimer with no glycans (yellow).

The ten sites with the largest median change in SASA or AbASA were plotted for each trimer in order to compare how the different glycoforms affected accessible surface area, and which regions were most affected (Figure 16).

For both the SASA and AbASA measurements, the heavily glycosylated V1/V2 and V4 regions were most prominently featured in the group of sites which had the highest change in their accessible surface area due to the addition of glycans (Figure 16A and B). The same sites repeated across protomers and were similar between the two trimers.

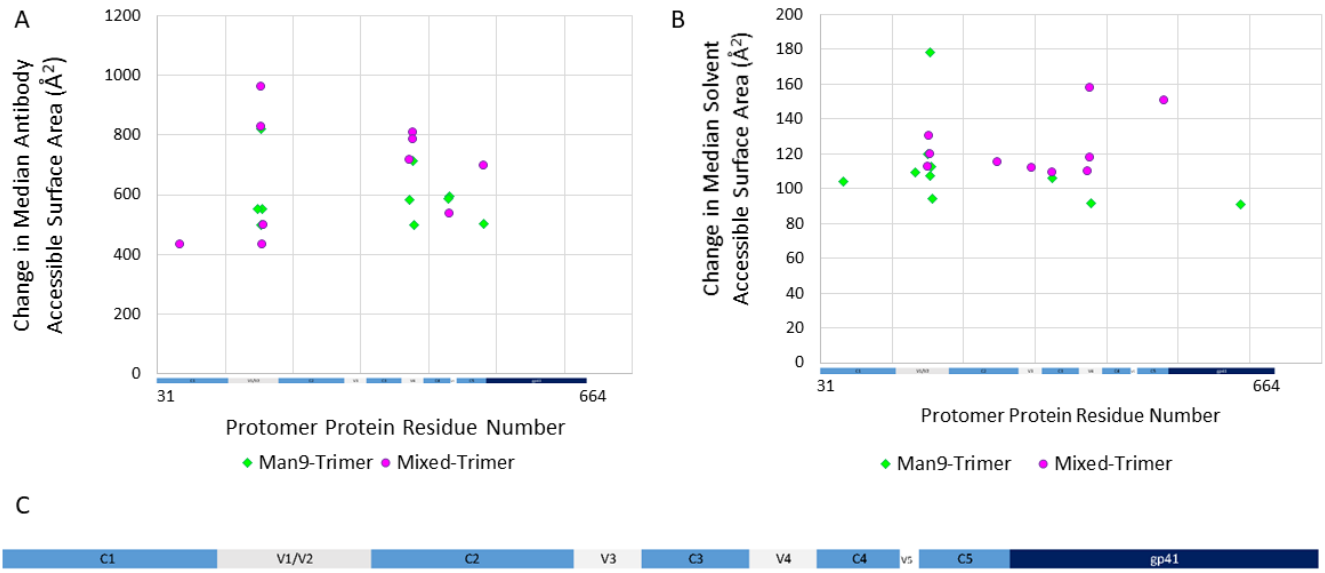


Figure 16: Solvent accessible and antibody accessible surface area of protein residues

A and B: Graphs of top ten sites of (A) ΔAbASA and (B) ΔSASA with the addition of glycans, for the two trimers. The majority of sites fall into the V1/V2 and V4 regions. C: Magnification of HIV-1 Env regions on x-axes of graphs A and B.

In order to locate these sites in context of the HIV-1 Env trimer, the top ten sites of accessible surface area change were plotted on the Mixed-Trimer and Man9-Trimer for both the AbASA (Figure 17) and the SASA results (Figure 18).

For both the Mixed-Trimer and the Man9-Trimer the top ten sites of AbASA change fell into the gp120 subunits of the trimers. Some of these sites were occupied by glycans: for the Man9-Trimer, these were site 411 on Protomer 2, and sites 185h and 462 on Protomer 3; and for the Mixed-Trimer these were sites 185h on Protomer 2 and 462 on Protomer 3 – both of which were occupied by a Complex-glycan. Notably, site 462 on Protomer 3 was included for both trimers. The majority of sites of highest AbASA change were located towards the apex of the trimers (Figure 17).

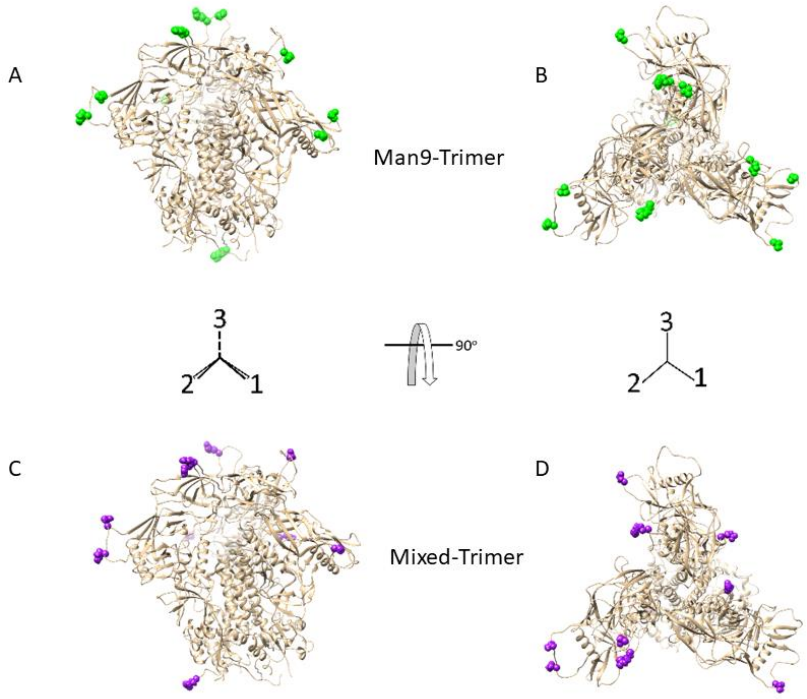


Figure 17: Sites of highest change in AbASA on the Man9-Trimer and Mixed-Trimer

The ten sites with the highest ΔAbASA were plotted onto the trimers for structural context. A and C: Side view. B and D: Top view of the Man9-Trimer and Mixed-Trimer.

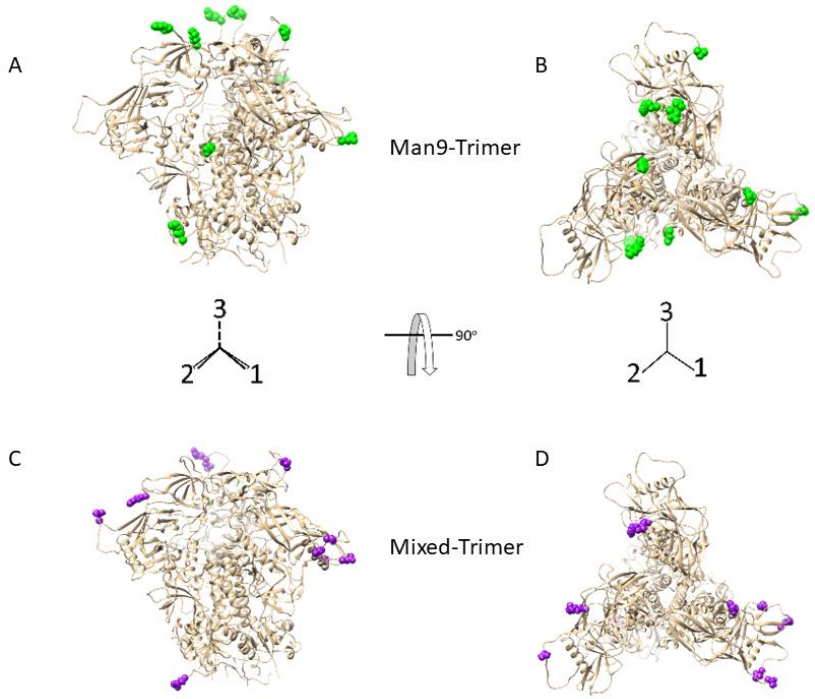


Figure 18: Sites of highest SASA change on the Man9-Trimer and Mixed-Trimer

The ten sites with the highest ΔSASA were plotted onto the trimers. A and C: Side view. B and D: Top view of the Man9-Trimer and Mixed-Trimer.

Similarly, the majority of the sites with the highest change in SASA were found at the apex of the two trimers (Figure 18). However, unlike the sites identified in the AbASA analysis, the spread of these sites was distinct between the Man9-Trimer and the Mixed-Trimer: sites were located close to the central axis of the Man9-Trimer, largely in the V1/V2 region (Figure 18B), and distal from the central axis of the Mixed-Trimer (Figure 18D).

The highest SASA change at sites where glycans were present, was at site 355 on Protomer 1, and sites 185h and 411 on Protomer 3 of the Man9-Trimer; and sites 355 and 406 on Protomer 1, and 185e on Protomer 3 of the Mixed-Trimer. Site 355 was common to both trimers, and was occupied by a Man5 glycan in the Mixed-Trimer.

Only 6 of the 20 amino acids were featured in the top ten sites of SASA or AbASA change (Table 8); interestingly all of these were polar amino acids. Notably, of the ten sites with the highest ΔSASA and ΔAbASA, six were located within PNGSs for the Man9-Trimer, and seven were located within PNGSs on the Mixed-Trimer (Table 8).

Table 8: Amino acids featured in the top 10 sites of SASA and AbASA change

		Man9-Trimer		Mixed-Trimer	
Amino acid	**Combined total**	**SASA**	**AbASA**	**SASA**	**AbASA**
Asparagine (N-glycan position)*	11	1_355_m9, 3_185h_m9, 3_411_m9	2_411_m9, 3_185h_m9, 3_462_m9	1_355_m9, 1_406_m9, 3_185e_c	1_185h_c, 3_462_c
Asparagine (X position)*	2	3_186	-	-	1_186
Arginine (X position)*	7	2_185f, 3_185f	2_185f, 3_185f	3_185f	2_185f, 3_185f
Arginine (non-PNGS)†	5	2_617	1_511	2_327, 3_511	3_511
Threonine (T position)*	6	-	1_408	1_278, 1_408, 2_408	1_408, 2_408
Threonine (non-PNGS) †	1	-	1_511	-	-
Glutamine	4	1_185c	2_403	1_185c	2_403
Glutamic Acid	3	1_62	1_185a	1_64	-
Lysine	1	2_171	-	-	-
PNGS total	26	6	6	7	7

* These amino acids are located within a PNGS, and their position refers to where they are found within the NXT/S motif.
† The classification "non-PNGS" is used where an amino acid is found both within and outside of a PNGS, and is included in order to clearly distinguish the residues that are not located within a PNGS.

Notably, although the third position of a PNGS can be occupied by either a threonine or a serine (NXT/S), no serine residues were included within the top sites of accessibility change, while seven threonine residues (six within PNGSs) were. This is particularly pronounced given that out of the 28 PNGSs per protomer sequence, 12 are NX**S** sites, and 16 are NX**T** sites.

In order to further investigate this, the PNGSs were grouped according to their motif (NXT or NXS), and the individual amino acid structures were further examined. At first glance, it would appear that threonine is found in the results more often because it is bigger than serine (93 $Å^3$ as compared to 73 $Å^3$) (Darby and Creighton (1993), cited in Simpson, 2003). However, on further inspection, it becomes apparent that the threonine residues with a large change in accessibility recur at site 408 (Table 8), and that it is thus the position of the residue, and not the type, that is significant. This site, examined in its three-dimensional context, proves to be on an exposed edge of the trimer, distal from the gp120 interfaces; interestingly, the size difference of the amino acids does appear to play a role here, since the threonine is adjacent to a serine on one side and a glycine on the other side – both relatively small residues – resulting in the threonine at this position being more exposed than the surrounding amino acids.

3.6. Investigating Enzyme and Solvent Accessible Surface Area as a Proxy for Glycan Processing

The secondary aim of this project was to investigate whether a computational approach to predicting the level of processing of a glycan can be used. Glycans are added to the nascent HIV-1 Env protein as $Glc_3Man_9GlcNAc_2$ precursors, and the first glycan in the assembly line is the Man9 glycan, which is formed by the removal of three glucose residues from the precursor. The Man9 glycan is then processed by the barrel-shaped enzyme, ER α-mannosidase I, which removes one mannose residue to form $Man_8GlcNAc_2$. Thereafter, the glycan undergoes further trimming and processing (See Section 1.4).

3.6.1. Enzyme accessible surface area

Recent work has highlighted the link between reduced glycan processing and lower solvent accessible surface area (Suga, Nagae and Yamaguchi, 2018); by the same token, the enzyme accessible surface area (EASA) was thought to have value in determining whether a Man9 glycan would be processed or not. In order to investigate this, Naccess (Hubbard and Thornton, 1993) was used to calculate the absolute EASA for every glycan residue. Since ER Mannosidase I is a barrel-shaped enzyme (Xiang, Karaveg and Moremen, 2016) and fits around a Man9 glycan in order to process it (Figure 19A), one

side of the barrel was measured to determine the required probe radius for Naccess (Figure 19B) (See Section 2.6).

Snapshots of the Man9-Trimer simulation were taken every 5 ns with CPPTRAJ (Roe and Cheatham, 2013), and the EASA values were determined (using a probe with radius 13 Å) for the terminal (0MA) glycan residues (Figure 19C).

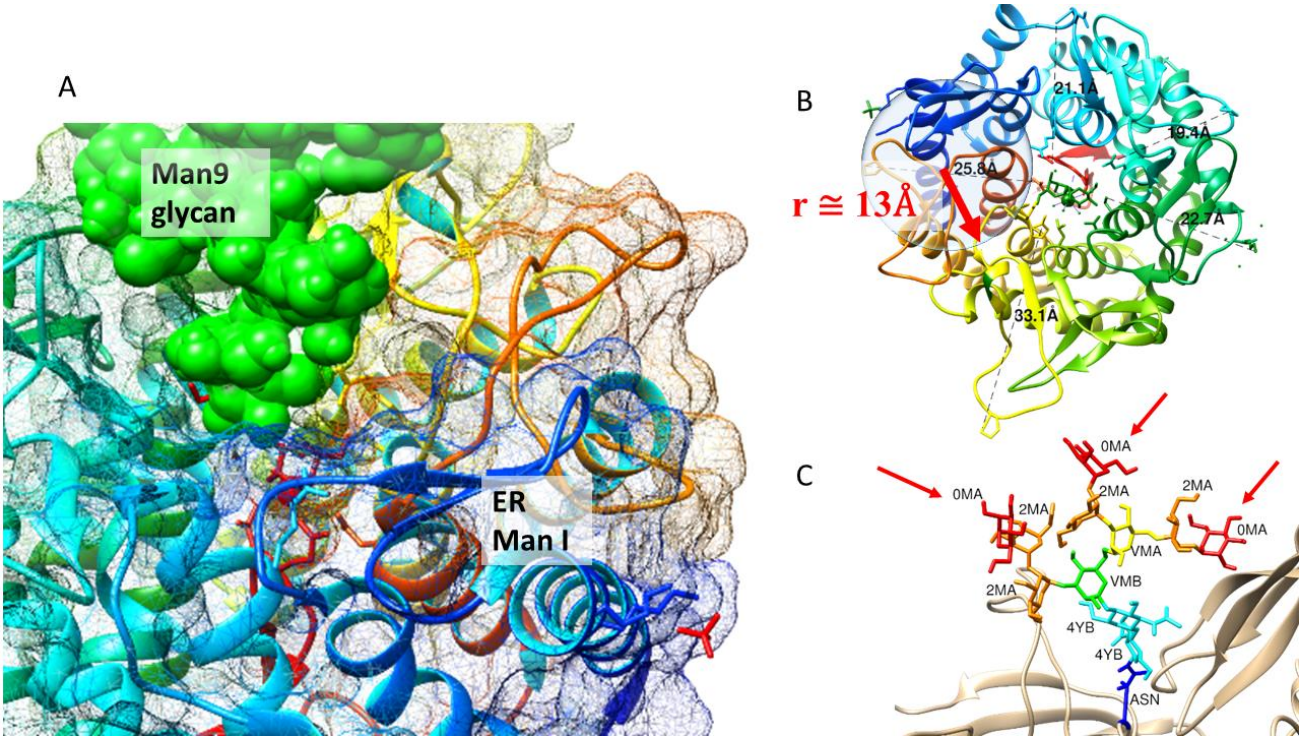


Figure 19: Basis of enzyme accessible surface area calculations for investigating processing prediction

A: Barrel structured ER Man I (cage and PDB 1X9D structure) around a Man9 glycan (green ball structure) for trimming of a terminal mannose residue. B: The 13Å radius (red arrow) of the Naccess probe was determined by measuring ER Mannosidase from the centre to the edges (PDB 1X9D (Karaveg *et al.*, 2005)). C: Man9 glycan showing its component residues. The terminal (0MA) residues (red arrows) were used in the accessible surface area calculations.

The absolute EASA measurements of the terminal glycan residues were binned according to the processed form of the corresponding glycans in the Mixed-Trimer, and the distributions were plotted in order to establish whether there is a relationship between these two variables (Figure 20).

The glycans that were allocated an unprocessed Man9 glycan form in the Mixed-Trimer had a higher EASA 52.3% more frequently than those that were given the intermediate Man5 glycan form. This was unexpected given that Man5 is more processed than Man9. Notably, the glycans that were processed to a Complex-glycan had higher EASA values 62.6% more frequently than the unprocessed Man9 glycans, and 64% more frequently than glycans that were intermediately processed to a Man5

glycan. This indicates that while the EASA may not be able to distinguish between high mannose forms, it may be a useful method for distinguishing between high mannose and complex forms.

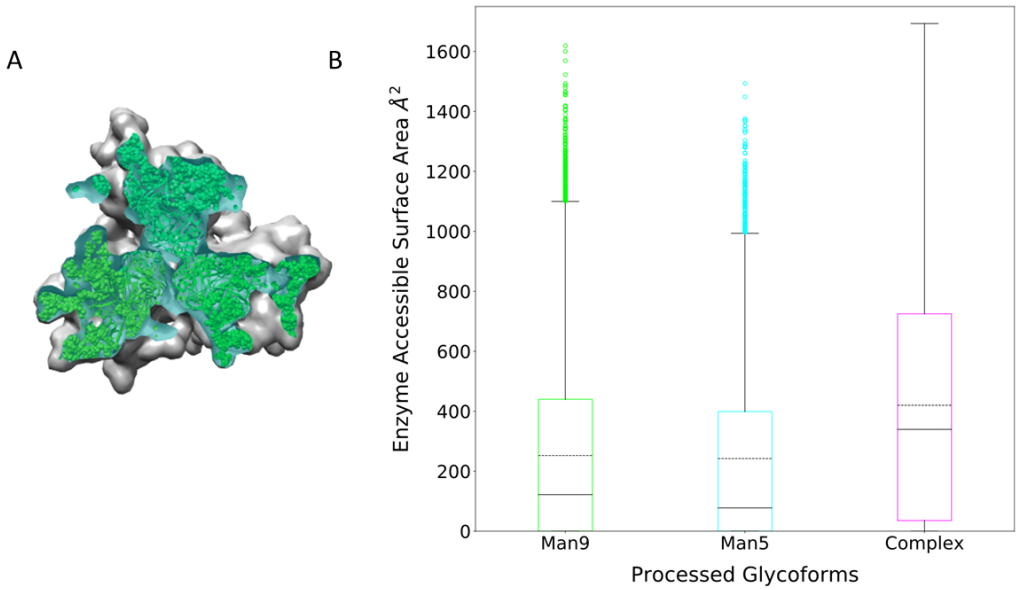


Figure 20: Investigating enzyme accessible surface area for processing prediction

A: 13 Å MSMS surface of Man9-Trimer showing z-slice through the trimer The Envelope protein is represented in green. Cyan is used to show the slice through the MSMS surface rendering, revealing the protein shape underneath. B: Boxplots of enzyme accessible surface area were plotted using the EASAs of the terminal (0MA) residues of each glycan on the Man9-Trimer, taken every 5 ns and grouped according to their corresponding processed glycoform on the Mixed-Trimer. Points depict outliers. Differences between Man9 and Man5 glycans, and Man9 and Complex-glycans were significant.

The median of the EASA values was then calculated for every terminal (0MA) glycan residue of the Man9-Trimer, and these values were summed and averaged to give an aggregate EASA value for every Man9 glycan. This aggregate was then converted to a percentage and plotted onto the trimer to visualise the accessibility in structural context (Figure 21 A and C). By comparing the results to the processing levels on the Mixed-Trimer (Figure 21 B and D), it can be seen that this aggregate measurement predicts a similar degree of glycan processing as would be present on the Mixed-Trimer, and potentially on the HIV-1 Env trimer as a whole, since the Mixed-Trimer represents one form of a glycosylated Env trimer.

Notably, both the intrinsic mannose patch around glycan 332, and the trimer-associated mannose patch at the protomer interfaces (See Section 1.4), are present on the EASA-predicted Man9-Trimer (Figure 21).

In contrast, while some of the complex glycans are correlated between the EASA-predicted Man9-Trimer and the Mixed-Trimer, such as glycans 618 and 462, others are not comparably represented, such as the 185e glycans (Figure 21).

Thus, since the glycan processing levels on the EASA-predicted Man9-Trimer and the Mixed-Trimer (used as a representative form of an HIV-1 Env trimer) were similar, it can be concluded that the EASA may provide a viable preliminary approach for determining the level of glycan processing present on the Env trimer.

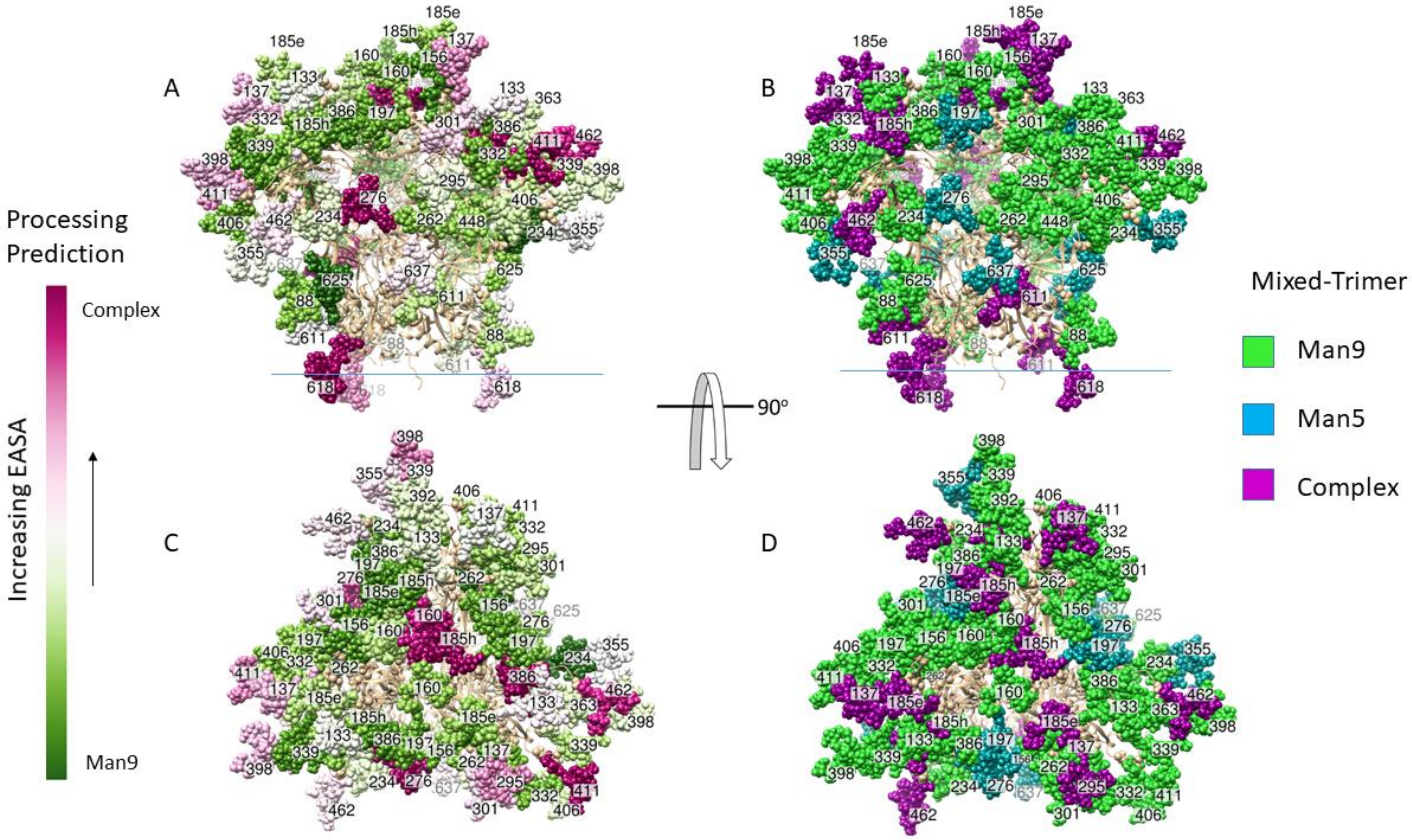


Figure 21: Comparison of EASA prediction with Mixed-Trimer glycoforoms

The EASA was explored as a proxy for predicting glycan processing. A and C: Man9-Trimer coloured by the average of the median EASA values for each glycan. For the "Increasing EASA" scale, a diverging green and purple colour scheme was used to indicate whether a glycan would be more likely to fall into the high mannose (green) or complex (purple) category, according to their EASA. B and D: The Mixed Trimer was included side-by-side with the plotted EASA values on the Man9-Trimer (A and C) in order to compare the results with what has been observed previously.

3.6.2. Solvent accessible surface area

In addition to the EASA, the solvent accessible surface area (SASA) was investigated in the context of glycan processing. SASA is a more commonly used metric and has been linked to the level of glycan processing previously (Suga, Nagae and Yamaguchi, 2018), and was thus interesting to compare with the EASA values.

Similar to the EASA analysis, z-slices were taken through the Man9-Trimer (Figure 22A), and using a 1.4 Å probe, the SASA was calculated every 5 ns. Each SASA value was then assigned according to the processed glycoform on the Mixed-Trimer (Figure 22B) (See Section 2.6).

The Man5 glycan SASA values were higher 54.5% more frequently than the Man9 SASA values, while the Complex-glycans had higher SASA values 56.4% more frequently than the Man9 glycans, and 51.7% more frequently than the Man5 glycans.

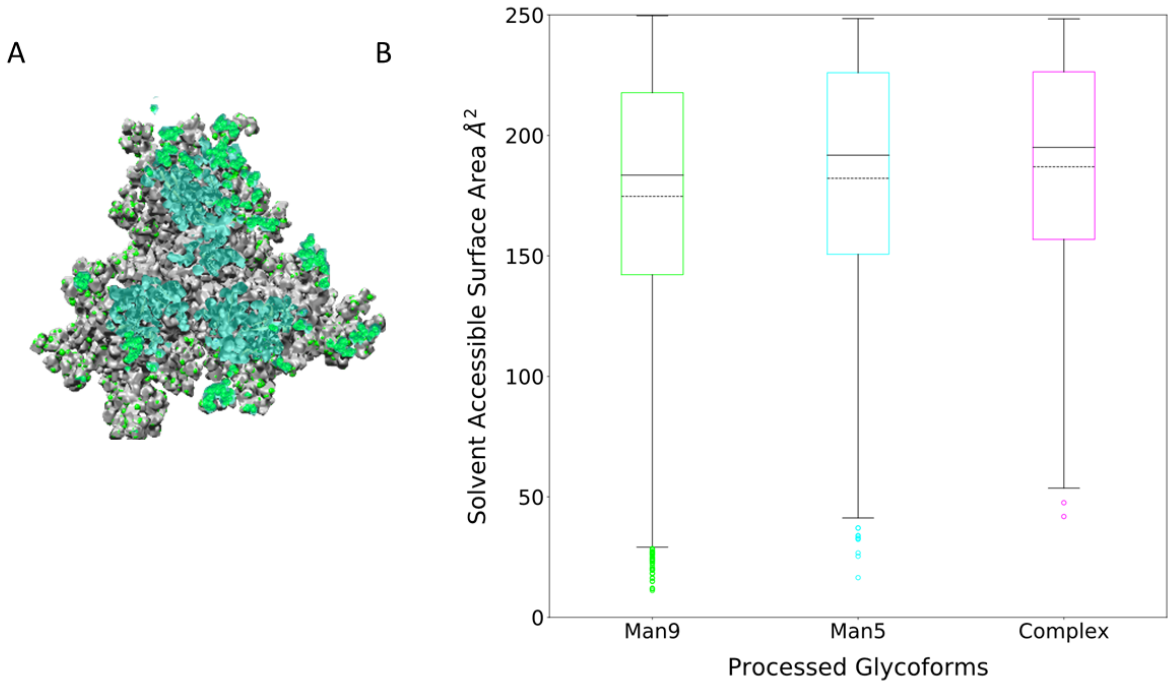


Figure 22: Investigating solvent accessible surface area as a proxy for glycan processing prediction

A: 1.4 Å MSMS surface of Man9-Trimer showing z-slice through the trimer. The Envelope protein is represented in green. Cyan is used to show the slice through the MSMS surface rendering, revealing the protein shape underneath. B: Boxplots of solvent accessible surface area were plotted using Naccess scores from the terminal (0MA) residues of each glycan on the Man9-Trimer, taken every 5 ns and grouped according to their corresponding processed glycoform on the Mixed-Trimer. Points depict outliers. The differences between the Man9 and Man5 distributions, and the Man9 and Complex-glycan distributions were significant, despite a clear overlap of values.

In order to compare the results from the SASA analysis to those of the EASA analysis, the medians were calculated for every 0MA residue and averaged. In order to create a finer gradient for the SASA values, since these were located within a narrower range than the EASA values were, the lowest of the average SASA values was subtracted from each average SASA value prior to their conversion to a percentile for plotting onto the trimer. The SASA-predicted trimer was compared both to the EASA-predicted trimer and the Mixed-Trimer (Figure 23).

In contrast to the EASA-predicted Trimer, the level of processing that was predicted by the SASA values correlated to the processing on the Mixed-Trimer to a much lower extent. Notably, although the glycan at position 332 was indicated as a high mannose glycan, overall the intrinsic mannose patch and the trimer-associated mannose patch, which are conserved across strains (See Section 1.4), were not apparent on the SASA-predicted trimer (Figure 23). It is possible that, because the SASA probe was a magnitude smaller and, therefore, a higher degree of accessibility occurred overall, this measure had a lower predictive power than that of the EASA.

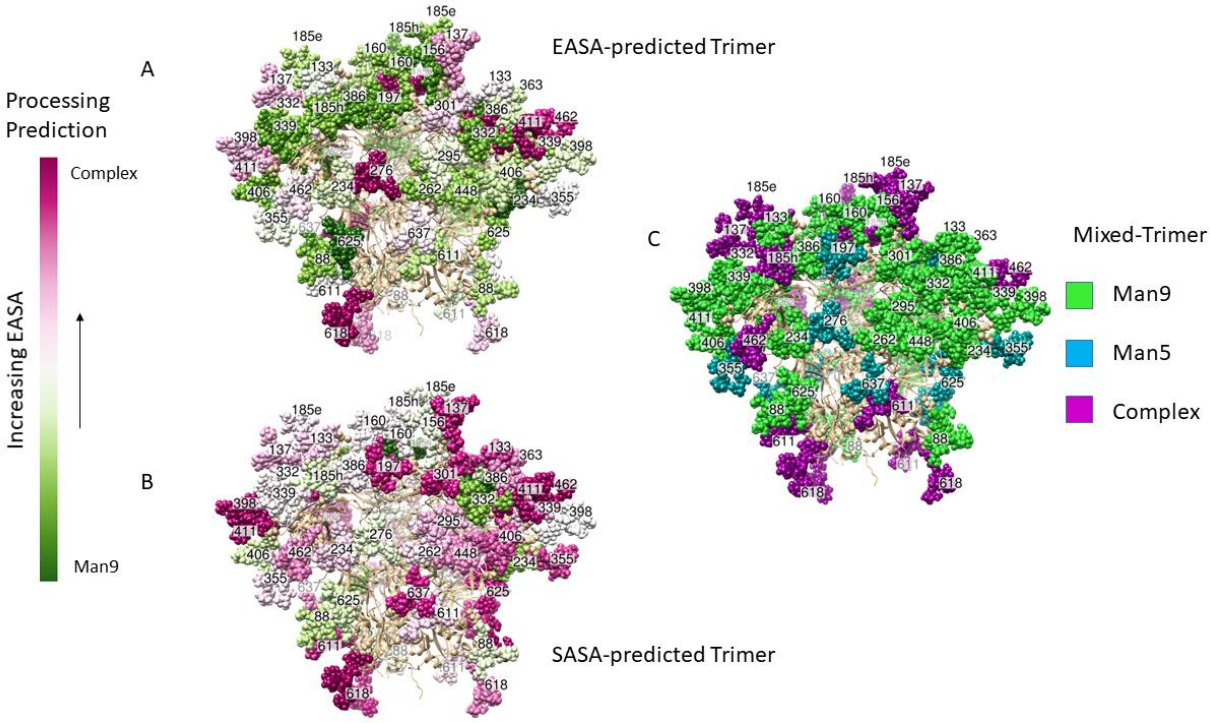


Figure 23: Comparison of processing levels of SASA-predicted Trimer, EASA-predicted Trimer, and Mixed-Trimer

The SASA was compared to the EASA as a proxy for predicting glycan processing. A: EASA-predicted Trimer. B: SASA-predicted Trimer. A diverging green and purple colour scheme was used to indicate whether a glycan would be more likely to fall into the high mannose (green) or complex (purple) category, according to their SASA or EASA. C: The Mixed-Trimer was included in order to compare the predictions to a laboratory-derived processing, as well as to each other.

4. Discussion

The picture alone, without the written word, leaves half the story untold ~ James Lafferty

HIV-1 is a global pandemic, and the virus is able to evade the immune system as well as multiple treatments by its rapid evolution, aided by the presence of the glycan shield that is put in place by host machinery (Eigen and Schuster, 1977; Wei *et al.*, 2003). These factors have contributed to a number of failed vaccine-strategy attempts (Klein *et al.*, 2013), and while HIV-1 is no longer the life sentence it once was, there is still no efficacious vaccine, and there are numerous side effects associated with ARVs. Additionally, ARVs are not always available, especially in the developing world. Thus, many believe that the way forward is to design an effective HIV-1 vaccine (Atun *et al.*, 2016; Stephenson and Barouch, 2016).

We have come closer to this goal with the discovery of broadly neutralising antibodies (Walker *et al.*, 2009), and the realisation that even HIV-1 has characteristic patterns that feature on its glycan shield, forming targets for these antibodies (Doores *et al.*, 2010; Pritchard, Vasiljevic, *et al.*, 2015). While there have been multiple structural studies over the years, the difficulty of studying glycans experimentally, combined with a lack of knowledge of the importance of glycans, resulted in the glycan shield being largely overlooked initially (Kwong, 1999; Chang *et al.*, 2007; Julien, Lee, *et al.*, 2013). With the knowledge that glycans form a part of most bNAb epitopes (Stewart-Jones *et al.*, 2016), there is a renewed push towards elucidating the structure and impact of the glycan shield in order to design an effective HIV-1 vaccine.

This thesis adds to the body of work on the HIV-1 glycan shield. Molecular dynamics simulations (500 ns in duration) were carried out on two HIV-1 Env trimer models that differed only in their glycan shields: a Man9-Trimer (including only Man9 glycans) and a Mixed-Trimer (including Man9, Man5, and Complex-glycans). The glycan distribution of the Mixed-Trimer was defined using three recent publications, which outlined mass spectrometry work on determining the glycan profiles of HIV-1 Env clones (Behrens *et al.*, 2016; Panico *et al.*, 2016; Cao *et al.*, 2017), in order to generate a biologically relevant model that resembled an HIV-1 Env trimer described in the laboratory (Figure 5C).

Overall, the results (Chapter 3) indicated that differences are indeed present due to the distinct glycoforms on these two trimers, and model-specific changes in accessible surface area of Man9 glycans, that are later processed, point to a rudimentary processing prediction method. This chapter,

therefore, moves on to discuss these findings in detail, outlining the results in light of the aims of this project, and setting the work against a backdrop of the body of literature.

4.1. The Influence of Different Glycoforms

The primary aim of this project was to investigate how different types of glycans influence interactions with the protein and surrounding glycans by comparing two computationally generated trimer models where one model was uniformly glycosylated with Man9 glycans and the other with a mix of glycans (Man9, Man5 or Complex-glycans; see Section 2.2). In this section, the impact of the different glycoforms on glycan occupancy, RMSD, RMSF, hydrogen bonding, and solvent and antibody accessible surface area will be discussed.

4.1.1. Glycoforms influence the level of glycan occupancy due to spatial constraints

It is well-known that spatial constraints hinder glycan processing in physiological conditions (Doores *et al.*, 2010; Pritchard, Vasiljevic, *et al.*, 2015). In this study, spatial considerations introduced by different glycoforms, i.e. the size and branching of different glycans, were also seen to affect glycan occupancy for computational glycosylation. Since glycans are added and processed as the protein is being made (Moremen, Tiemeyer and Nairn, 2012), and glycan occupancy has been seen to differ in mass spectrometry experiments both on trimers (Cao *et al.*, 2017, 2018) and gp120 monomers (Yu *et al.*, 2018), this may have relevance for biological conditions, as well as computational glycosylation algorithms.

Differences in glycan occupation between the ten Man9-Trimers and the Mixed-Trimers occurred at two sites. Site 185h, which is situated on the V2 loop, had a noticeably lower occupancy on the Mixed-Trimers; and site 392, which is situated on the V4 loop, had a considerably lower occupancy on the Man9-Trimers (Figure 6A and B). The V1/V2 and V4 loops are both densely glycosylated regions of HIV-1 Env (Figure 5); thus it is unsurprising that there appear to be spatial constraints for the glycans in these regions.

Site 185h, which had lower occupancy on the Mixed-trimers (Figure 6A), is located at the apex of the trimer (Figure 6B) and surrounded by glycans from the V1/V2 region, amounting to a dense cluster of nine Complex-glycans on the trimer apex. Although incomplete on the Mixed-trimers, the computational glycosylation process successfully glycosylated each of these sites on the Man9-Trimers. One explanation for this may be due to the structures of the glycans themselves; the Complex-glycan used in this study has an extra sugar residue (See Section 2.2), which contributes to a longer

glycan structure compared to a Man9 glycan, thereby limiting the size of the Complex-glycan rich cluster.

Cao *et al*. (2017) also found reduced glycan occupancy at site 185h, as well as 185e of the BG505.SOSIP trimer, with 89% and 82% occupancy respectively. Although the results from Yu *et al*. (2018) are based on a gp120 monomer, they include 99 strains of HIV-1, and these results clearly illustrate the differing occupancy of PNGSs on Env. Site 185h was not included in the study by Yu *et al.* (2018), but nearby sites on the V1/V2 loop exhibit incomplete glycosylation, averaging 70% occupancy.

In contrast to site 185h at the apex, site 392 is located towards the outside of the lobes of gp120, near the CD4 binding site (Figure 6B), and had a considerably lower occupancy on the Man9-Trimers (Figure 6A). Interestingly, the glycan at site 392 takes on a Man9 glycoform in both trimers. However, other glycans that cluster nearby on the folded protein are different glycoforms on the two trimers, and it is likely that this causes the spatial constraint that results in occupancy differences. On the Man9-Trimers, glycans 197 and 276, which also cluster around the CD4 binding site, are Man9 glycans. However, on the Mixed-Trimers, these are Man5 glycans and thus are shorter and take up less space (Figure 5) – leaving more room for the addition of glycan 392 in this region.

Although variable occupancy was not observed in the experiments by Cao *et al.* (2017, 2018), the average occupancy across strains for site 392 in the experiment by Yu *et al.* (2018) was 84%, highlighting the variable occupancy of this site.

Interestingly, sites 618 and 625 had greatly reduced occupancy (63.8% and 62% respectively) in the study by Cao *et al.* (2017), which was not reflected in the computationally glycosylated models. This may be due to the presence of a membrane when glycosylation takes place *in vivo* (Checkley, Luttge and Freed, 2011), as opposed to the computational models that do not contain a membrane, and thus have more space available for the addition of glycans at these sites. Thus, future experiments should perhaps take into account the steric hindrance that is present at membrane-proximal sites.

Thus, we can see that the presence of different glycoforms does indeed impact the glycan occupancy at various sites – and importantly, not just sites that are located in close proximity on the linear sequence, but also sites that are located close together in three-dimensional space, since a larger glycan can be accommodated if there is a small glycan on a nearby site.

The final Man9-Trimer did not include a glycan at site 363 on Protomers 1 and 2, and at site 392 on Protomers 2 and 3; while the final Mixed-Trimer lacked glycans at sites 386 and 462 on Protomer 1,

185h and 392 on Protomer 2, and 363 on Protomer 3 (Table 3). Notably, each of these sites exhibited variable occupancy across HIV-1 Envelopes in the study by Yu *et al.* (2018), highlighting that incomplete glycosylation occurs in nature, and supporting the validity of these models as possible forms of Env.

4.1.2. The Mixed-Trimer is more stable than the Man9-Trimer

As well as affecting the glycan occupancy, different glycoforms also impacted the stability of the trimers. The RMSD and RMSF data highlight the greater stability of the Mixed-Trimer, which generally achieved lower deviations and fluctuations from the reference structure than those of the Man9-Trimer (Figure 7 and Figure 8).

The higher levels of hydrogen bonding on the Mixed-Trimer (1 332 406, compared with 1 067 251 for the Man9-Trimer) may contribute to this difference in stability. The only difference between the two trimers was their glycan shields and, therefore, the impact of different glycoforms on hydrogen bonding was examined, since this has been shown to be important to the integrity of the glycan shield (Stewart-Jones *et al.*, 2016).

Interestingly, Man5 and Man9 glycans at equivalent positions between the two trimers contributed to the longevity of hydrogen bonding to a similar degree; however, the difference between equivalent Complex-glycans and Man9 glycans was significant, and Complex-glycans had a 29.7% higher median presence in the hydrogen bonds over time. (Table 7).

The lack of significant differences between the hydrogen bonding involvement of equivalently positioned Man5 and Man9 glycans may be due to the similarity in branching structure of the Man5 and Man9 glycans (See Section 2.2), as well as their location in glycan dense regions. In contrast, the Complex-glycans could extend beyond the rest of the glycan shield due to their greater length. Additionally, the Complex-glycans were frequently located in glycan-sparse regions, allowing for interaction with glycans located further away due to the unrestricted movement from their attachment point, rather than having to navigate a densely glycosylated space.

Interestingly, another study (Gift *et al.*, 2017) identifies complex glycans as reducing stability by comparing the thermal stability of viral Envs produced in wild-type HEK-293T cells to that of those produced in the presence of kifunensine (which would result in uniformly Man9 glycan shields, such as with the Man9-Trimer). This is in contrast to what was observed here, where the more biologically relevant Mixed-Trimer, which includes complex glycans, is more stable than the Man9-Trimer. Notably, while molecular dynamics has been used to explore factors that may influence thermal stability (Gu *et al.*, 2018), the direct relationship between thermal stability and the RMSD (used as a

measure of stability in MD) of a molecule during a simulation has not been investigated. While more work may be necessary to understand these differences, it is clear that trimer stability is important for the elicitation of neutralising antibodies (Feng *et al.*, 2016).

4.1.3. The presence of complex glycans at the HIV-1 Env apex may be important for interprotomer stabilisation

A large proportion of Complex-glycans were present on the apex of the Mixed-Trimer on the V1/V2 loop, and notably this is where the majority of its 177 778 interprotomer stabilising hydrogen bonds occurred (Figure 10B and D). Laboratory studies have shown that trimers which have been chemically cross-linked between their gp120 monomers at the trimer apex, using glutaraldehyde, have increased stability and antigenicity (Schiffner *et al.*, 2018), and it is plausible that the interprotomer hydrogen bonds between Complex-glycans at the trimer apex in this study fill the same role.

In contrast, although both trimers exhibited their highest levels of hydrogen bonding at the trimer apex, the 96 551 Man9-Trimer interprotomer hydrogen bonds occurred largely around the sides of the trimer, with only one notable apex-located interprotomer interaction (25 757 hydrogen bonds) between glycans 185h on Protomer 1 and glycan 160 on Protomer 3 (Figure 9B).

The highly conserved glycan at site 197 (Travers, 2012), which is important to neutralisation resistance in multiple HIV-1 strains (Wyatt *et al.*, 1995; Ly and Stamatatos, 2000; Li *et al.*, 2008; Utachee *et al.*, 2010; O'Rourke *et al.*, 2012), was involved in the other 67.5% of the dominant interprotomer interactions on the Man9-Trimer.

Complex-glycans were present in 83.5% of the dominant interprotomer hydrogen bonds on the Mixed-Trimer. The equivalently positioned glycans on the Man9-Trimer were only involved in 32.5% of the dominant interprotomer hydrogen bonds. As before, this may be due to the extension of the Complex-glycans beyond the rest of the glycan shield, as well as their added length compared to the Man9 glycans, allowing the Complex-glycans to form long-distance interactions with glycans on other protomers, and contributing to increased stability of the Mixed-Trimer compared to the Man9-Trimer.

The interprotomer interactions on the HIV-1 Env trimer have not been closely investigated, although they are referred to in passing (for example, in Stewart-Jones *et al.* (2016)); and some interactions, such as between the V1/V2 stem and the tip of the neighbouring protomer V3 loop, similar to the glycan 197 interactions on the Man9-Trimer, have been described (Li *et al.*, 2008). Thus, this is an area that could benefit from further research, especially given the evidence that complex glycans at the

apex may contribute to trimer stability and antigenicity, which are important for the design of vaccine immunogens.

4.1.4. Scissoring and trimer asymmetry in Man9-Trimer likely caused by lack of stabilising complex glycans

Interestingly, Protomers 1 and 3, as well as Protomers 2 and 3, interact, but there is no notable interprotomer interaction present between Protomers 1 and 2 of the Man9-Trimer. This may contribute to the scissoring of protomers that was observed in the simulation of the Man9-Trimer, and which resulted in an exposed CD4 binding site between Protomers 1 and 2 (Figure 12).

This is similar to the results described in the study by Lemmin *et al.* (2017), where MD simulations were carried out using a uniform Man5 glycan shield on the clade A BG505 SOSIP.664 trimer. In both the current study, and the study by Lemmin *et al.* (2017), one of the three CD4 binding sites was exposed, while access to the other two was restricted. Interestingly, scissoring was not observed in the study by Ferreira *et al.* (2018), which used a uniformly glycosylated Man9 shield, similar to the Man9-Trimer here. However, Ferreira *et al.* (2018) used a clade C strain (CAP45.G3) of HIV-1, rather than the clade A BG505 SOSIP.664 trimer that was used in this study, and it is possible that differences in the distribution of PNGSs between the sequences may have contributed to the contrasting observations.

The Mixed-Trimer, unlike the Man9-Trimer, did not exhibit asymmetry in its interprotomer hydrogen bonding, and limited scissoring was observed (See Section 3.4.3). The strong 3-way interprotomer stabilisation on the Mixed-Trimer may prevent the occurrence of scissoring.

Consequently, it would appear that complex glycans, including the Complex-glycans used to glycosylate the Mixed-Trimer in this study, and which the Man9-Trimer lacks, are important to forming the stabilising hydrogen bonds between the protomers on the trimer apex. However, a different underlying protein structure, where the location of PNGSs and Man9 glycans allow interprotomer interactions, may contribute to the lack of scissoring observed for other Man9-glycosylated trimer models such as in the study by Ferreira *et al.* (2018).

These differences are important to note if computational methods are to be used in antibody studies. For example, the prediction of the degree of interaction between the CD4 binding site antibody, VRC01 (Wu *et al.*, 2010), is likely to be different for the Man9-Trimer, which exhibited scissoring, as compared to the Mixed-Trimer, which did not.

It is worth noting that the study by Lemmin *et al.* (2017) reported the observation of scissoring in the 0.5 – 1.0 μs range of the simulation. Since one of the caveats of molecular dynamics relates to the

length of the simulation, where all conformations are known to have been sampled, it could be speculated that perhaps the Mixed-Trimer would exhibit scissoring if the 500ns simulation used here had been continued for a longer time span, or if a replica exchange method (which samples more states) had been used, such as in the study by Yang *et al.* (2017). However, it could also be argued that the study by Lemmin *et al.* (2017) did not include complex glycans, which may have negated the scissoring effect.

4.1.5. The 185e glycan compensated for an absent 185h glycan, highlighting the importance of interactions at the trimer apex

The importance of the Complex-glycans to interprotomer interactions is highlighted by the Complex 185e glycan filling the role of the absent 185h Complex-glycan on Protomer 2 of the Mixed-Trimer. On Protomers 1 and 3, glycan 185e played a negligible role in interprotomer hydrogen bonding interactions while glycan 185h featured in the dominant interprotomer interactions. However, on Protomer 2, which lacks the 185h glycan, glycan 185e was involved in the dominant interprotomer interactions (Figure 11 and Table 5). This points to structural rearrangements within the glycan shield, such as those observed in Ferreira *et al.* (2018), taking place to maintain the integrity of the glycan shield, and the inclusion of interprotomer interactions in this rearrangement is remarkable.

4.1.6. Hydrogen bonding is more dispersed on the Mixed-Trimer than on the Man9-Trimer

In addition to the stabilisation caused by the situation of Complex-glycans at the trimer apex, the greater dispersion of glycan-glycan hydrogen bonds over the Mixed-Trimer as compared to the Man9-Trimer may also increase its stability.

The Mixed-Trimer had 8% more unique glycan pairs that formed hydrogen bonds compared to the Man9-Trimer. While some asymmetry was seen in the favouring of Protomer 1 glycans in its hydrogen bonding (Figure 9B), overall the Mixed-Trimer's hydrogen bonds were more dispersed over the surface of the trimer than was seen on the Man9-Trimer. Additionally, the Mixed-Trimer had 13 glycans that were predominantly present in its set of glycans involved in hydrogen bonding, compared to 7 on the Man9-Trimer. The larger group of predominantly present glycans observed for the Mixed-Trimer may have contributed to the better compensation on the Mixed-Trimer when one glycan is removed, such as with glycans 185e and 185h (Section 4.1.5), where the hydrogen bonds were redistributed.

Additionally, the greater dispersion of hydrogen bonds on the Mixed-Trimer may suggest a less easily penetrated glycan shield: Stewart-Jones *et al.* (2016) point out the importance of glycan-glycan interactions, and notably hydrogen bonds, to maintaining the integrity of the glycan shield. When the

glycan shield acts as a barrier to antibodies (Wei *et al.*, 2003), greater dispersal of hydrogen bonds may help to prevent penetration of the glycan shield by these antibodies at areas of weaker bonding. On the other hand, when glycans form part of an antibody's epitope (Gray *et al.*, 2011; Walker *et al.*, 2011), more closely bonded glycans over the glycan shield may impede the binding of these antibodies.

Although differences in the dispersion of hydrogen bonds were observed between the trimers, one commonality was that the hydrogen bonds were unevenly distributed across protomers (Figure 9 A and B). While more sampling may have resulted in evenly dispersed hydrogen bonding, the uneven spread of hydrogen bonding may be valid in a biological context if the asymmetry occurs during the time at which an antibody is binding. Notably, some antibodies, such as PG9, interact asymmetrically with HIV-1 Env (Julien, Lee, *et al.*, 2013), and thus the asymmetry in hydrogen bonding could be beneficial in such cases.

While much focus was placed on the hydrogen bonding dynamics of the two trimers in this study, as these have been shown to be important to the glycan shield (Stewart-Jones *et al.*, 2016), this does not rule out the occurrence of other interactions, such as Van der Waals, and hydrophilic interactions, which could be addressed in future studies.

4.1.7. Areas around glycans, including important epitopes, exhibit the largest change in accessibility, regardless of glycoform

While glycoforms were important to the dynamics of the HIV-1 Env trimers in this study, it was not the type of glycan as much as their presence or absence that most affected the accessible surface area of protein residues. This was shown by the SASA and AbASA analysis, where the change in accessibility on removal of glycans was determined for the Man9-Trimer and Mixed-Trimer at every protein residue (See Section 3.5).

Interestingly, the residues included in the top ten sites of change in accessibility (upon glycan removal), predominantly occurred in the V1/V2 and V4 loops (Figure 16A and B). This may be due to the exposure of these regions compared to other regions of HIV-1 Env, as well as the high density of PNGSs (contributing to a high level of shielding when glycans are present).

Strikingly, out of the 20 amino acids, only 6 were featured in the top ten sites of median SASA or AbASA change (Table 8), and all were polar amino acids. This may be explained by a preferential interaction of glycans, which are highly polar by nature, with polar amino acids; this interaction would reduce accessibility of these sites in the presence of glycans.

Additionally, of the top sites of median SASA or AbASA, 60% were within a PNGS on the Man9-Trimer, and 70% were within a PNGS on the Mixed-Trimer (Table 8), highlighting the role of glycan proximity in determining the accessibility of a protein residue.

This is important in the context of antibody binding, since the sites that are more accessible in the absence of a glycan, such as sites 411, 462 and 185h as highlighted as in the antibody accessible surface area analysis (Table 8), may have important roles in exposing antibody epitopes. For example, the removal of glycan 185h would create a glycan hole at the apex of the trimer, which may aid the PGT145 antibody (Walker *et al.*, 2011) in extending into the junction of the three protomers to contact the underlying protein epitope (Liu *et al.*, 2017). Similarly, gp120 sites 455 - 465 are important for VRC01 binding (Zhou *et al.*, 2010), and thus the removal of the glycan at site 462 may improve the binding of VRC01 by exposing this region.

4.1.8. Repeated sites of high accessibility change may offer targets for creating glycan holes

Notably, a number of the sites with high levels of change in accessibility due to the presence of a glycan were repeated between the two trimers (Table 8). If this observation is consistent across a wide variety of models, then these sites may be useful targets for creating glycan "holes" in the interest of designing immunogens for the priming step of a prime-and-boost vaccine strategy.

Glycan holes have been explored in a number of studies – from observing those that occur naturally when a glycan is absent, such as the hole created by a missing 197 glycan on a JR-FL trimer (Crooks *et al.*, 2015), or those generated by the absence of glycans at sites 241 and 289 on a BG505 SOSIP.664 trimer (McCoy *et al.*, 2016); to those created in the laboratory, as was carried out by Zhou *et al.* (2017).

In this latter study (Zhou *et al.*, 2017), four glycans were removed (by mutating their PNGS motifs) around the CD4 binding site of clade C Envelopes: glycans 197, 276, 363, and a glycan from the 460-463 region (notably, glycan 462 was found to be a site of high change in accessibility in the current study (Table 8)). Deglycosylation resulted in 10 to 100-fold increase in sensitivity to neutralisation at the CD4 binding site (Zhou *et al.*, 2017). Although not carried out in the study, it was proposed that the deglycosylated trimers could act as primers, and by restoring the removed glycans prior to the boosting step, antibodies may develop glycan tolerance either through direct engagement (Garces *et al.*, 2015) or by accommodating shifts in the glycan conformation (Stewart-Jones *et al.*, 2016), as has been observed in B-cell ontogeny studies.

Furthermore, glycan holes have been studied in the context of how complete (i.e. with a higher number of glycans, and fewer glycan holes) a glycan shield is at viral transmission (Wagh *et al.*, 2018), highlighting that more complete glycan shields are advantageous to the development of broadly

neutralising antibodies, and suggesting that the exclusion of glycan holes is important to the development of neutralisation breadth.

Thus, these repeated sites that impact the accessibility of the Env trimer may be worth exploring further in the context of developing bNAbs.

4.1.9. Further glycan heterogeneity may affect the dynamics of the glycan shield

It is apparent that, while the heterogeneity of the glycan shield of the Mixed-Trimer is a better representation of the HIV-1 Env than has previously been used in MS studies, it is still not fully representative of the diversity that occurs within circulating viral strains.

Tri- and tetra-antennary glycans are known to be found on the glycan shield (Crispin, Ward and Wilson, 2018); these glycans are larger than any of the glycans represented in this study (which used a bi-antennary Complex-glycan), and thus may affect the dynamics of the glycan shield in a different way.

The fucosylation of the glycan core has not been explored either, and whether this affects the dynamics of the glycan shield as a whole, or just the dynamics of the individual fucosylated glycans, is as yet unknown and may be important for those antibodies that contact the $Man_3GlcNAc_2$ core.

Additionally, while molecular dynamics is a good method for exploring glycan rotamers (Woods and Tessier, 2010), the 500 ns length of the simulation may not have been enough to fully sample all possible rotamers at every glycan position. While the scope of the current project did not allow for longer simulations to be generated prior to analysis, the conformational sampling could be improved by increasing the length of the study, or using replica exchange.

Finally, even within the set of glycans that was used, different linkages are possible for both the Man5 glycan and the Complex-glycan, and it remains to be seen to what extent these linkages affect the dynamics of the glycan shield.

4.1.10. Non-identical glycosylated PNGSs may influence the dynamics

It was noted previously that different glycoforms changed the levels of glycan occupancy, and also that differing glycan occupancy is to be expected across even clonal HIV-1 Env populations (See Section 4.1.14.1.14.1.14.1.1). While this underscores the applicability of the results, it still leaves the question of what differences might be observed were identical PNGSs glycosylated on both trimers.

The most notable implication is that two Complex-glycans were not added to the Mixed-Trimer at positions where Man9 glycans were added to the Man9-Trimer (positions 462 on Protomer 1, and 185h

on Protomer 2). This is a particularly interesting omission, since it reduces the number of Complex-glycans present on the Mixed-Trimer, and yet the Mixed-Trimer still has a higher level of hydrogen bonding and stability compared to the Man9-Trimer, highlighting the importance of Complex-glycans in these functions. Additionally, these PNGSs were among the highest sites of change in antibody accessibility when a glycan was removed (See Section 3.5).

The omission of the glycan at position 363 on Protomer 2 of the Man9-Trimer may have had some influence on the dynamics, but as this glycan (when present) did not form the predominantly present or absent hydrogen bonds, and it was not present in the set of PNGSs of largest change in solvent or antibody accessible surface area, the effects of its omission would likely have had a fairly low impact on the general dynamics of the Man9-Trimer.

The other omitted glycans on the two trimers map largely equivalently: both trimers lack the Man9 glycans at position 392 on Protomer 2, and 363 on Protomer 3; while the last missing Man9 glycan is 6 amino acids apart on Protomer 1 of the two trimers, at position 392 on the Man9-Trimer, and position 386 on the Mixed-Trimer. It is unlikely that these largely equivalent missing glycans would have resulted in significantly different dynamics between the two trimers, however, future MD studies will add further evidence to this assumption.

4.2. Potential for Glycan Processing

The secondary aim of this project was to estimate the potential for glycans to be further processed from the initial Man9 glycan. Here, as a preliminary investigation, this was done by carrying out enzyme and solvent accessibility analyses, since the physical space available to glycan-processing enzymes may influence the subsequent processing.

Since the intrinsic mannose patch and the trimer associated mannose patch occur in areas where there are spatial constraints for processing enzymes (Doores *et al.*, 2010; Pritchard, Vasiljevic, *et al.*, 2015) (See Section 1.4), areas of high accessibility were expected to correspond to more processed glycans (tending towards complex), and areas of low accessibility were expected to correspond to high mannose glycans, which are less processed.

4.2.1. The choice of ER ManI as opposed to other enzymes

The ER α-mannosidase I enzyme was used as the basis for the enzyme accessibility calculations, since this enzyme targets Man9 glycans, and the Man9-Trimer was used as the input for the Naccess (Hubbard and Thornton, 1993) calculations. Although the Golgi mannosidase I A – C enzymes also process high mannose glycans once Env has moved to the Golgi apparatus following oligomerisation (Doores *et al.*, 2010), these enzymes were not included in the preliminary investigation.

In the study by Lobsanov *et al.* (2002), it was shown that the ER and Golgi α-mannosidase I enzymes were structurally similar. On overlaying the *Penicillum citrinum* α-mannosidase I enzyme (which is in the same class as Golgi α-mannosidase I; PDB ID: 1KKT (Lobsanov *et al.*, 2002)) with the 1X9D structure (Karaveg *et al.*, 2005) used to determine the probe size for ER Man I, it was found that these enzymes would require the same probe size (data not shown) to measure the enzyme accessible surface area, and thus the use of only the ER α-mannosidase I enzyme in this preliminary study, was justified.

4.2.2. The impact of oligomerisation

Since trimming of Man9 glycans (from Man9 to $Man_8GlcNAc_2$) takes place while Env is in protomeric form in the ER, and then the majority of trimming takes place following trimerisation and translocation to the Golgi, the accessibility calculations were carried out on a trimer. However, it may be interesting to compare this to calculations performed on a protomer in a future study in order to understand how oligomerisation may affect processing prediction.

4.2.3. EASA may be useful for differentiating high mannose and complex glycans

With Complex-glycans exhibiting higher EASA measures 62.6% and 64% more frequently in comparison to Man9 and Man5 glycans respectively, the EASA calculations provide a rudimentary

method for predicting whether a glycan will be of a complex or high mannose form. The EASA distribution differences were significant, with p-values in the region of 1×10^{-100} when comparing Complex-glycans to either high mannose distribution.

The impact of this was clear when an aggregate measure of the accessibility of the terminal glycan sites was plotted onto the trimer and compared to the Mixed-Trimer: the intrinsic mannose patch (Doores *et al.*, 2010) and trimer associated mannose patch (Pritchard, Vasiljevic, *et al.*, 2015) (See Section 1.4) were readily identifiable according from the processing prediction results (Figure 21). Since snapshots were taken every 5 ns, giving a total of 90 snapshots, these results should reflect a representative sample of the accessibility.

4.2.4. Higher predictive power of the EASA compared to the SASA

Since the a higher solvent accessible surface area has previously been linked to increased glycan processing (Suga, Nagae and Yamaguchi, 2018), this was also measured for the terminal residues of every glycan. Interestingly, when normalised and compared to the EASA-predicted Man9-Trimer and the Mixed-Trimer in structural context, it became apparent that the SASA overrepresented the level of processing that would occur on the trimer (Figure 23). Thus, it appears that the EASA may provide a more accurate prediction of the glycan processing that might occur *in vivo* compared to the commonly used SASA measure.

4.2.5. EASA indicates potential for a higher number of complex glycans in the Mixed-Trimer

Unexpectedly, the Man9 glycans had a higher EASA than the Man5 glycans 52.3% more frequently; this was significant with a p-value of 2.04×10^{-5}. Here, however, it could be argued that in fact some of the Man9 glycans should have been allocated a complex glycoform, since Man9 was allocated as the default when glycan processing information was lacking or contradictory (See Section 2.2). This was also reflected in the structural plot of the EASA values, which showed potential for greater processing at a number of sites when compared with the Mixed-Trimer (Figure 21).

Indeed, in the findings of Cao *et al*. (2018), virion derived BG505 Env glycan shields had a higher proportion of complex glycans than was found for the recombinant BG505 SOSIP.664 constructs; this indicates that the Mixed-Trimer (which is based on BG505 SOSIP.664) may be able to accommodate more complex glycans than was assigned in this study.

4.2.6. Glycoforms chosen at variable and uncharacterised sites align well with EASA predictions

In cases where the information on the glycoform at any given site was variable, inadequate, or conflicting, the glycoform was assigned to represent the available information as closely as possible, or a Man9 glycan was added as the default (See Section 2.2).

Sites 197, 276 and 355 were assigned a Man5 glycan due to their intermediate levels of processing. Interestingly this is reflected on the EASA-predicted trimer: the 197 glycans are all slightly processed high mannose glycans, the 355 glycans fall into an intermediate category of processing and may form hybrid glycans, and the 276 glycans each take on a different category, with one high mannose, one intermediate-complex, and one complex (Figure 21). Site 406 was assigned a Man9 glycan based on incomplete and conflicting information for this site; all of the glycans at position 406 were predicted to be a slightly processed high mannose on the EASA-predicted trimer (Figure 21).

4.2.7. The effect of rotamers on accessibility may not be sufficiently represented

Since the simulations are unlikely to have sampled all possible conformations, it is likely that not all glycan rotamers were taken into account, which may have affected the processing results. However, since it is likely that not all conformations are sampled while glycans are being processed *in vivo*, these results represent one possible form of processing that could occur.

4.2.8. Processing prediction and glycan holes

The ability to predict the level of processing may be useful when designing glycan holes for vaccine immunogens, since the glycoforms may affect interactions with antibodies. For example, Behrens *et al.* (2018) noted only slight increases in the glycan processing of glycan 363, in spite of the deletion of the glycans at the neighbouring sites (197, 386, and 462). In contrast, Zhou *et al.* (2017) noted a change in processing from high mannose to complex glycans at positions surrounding the hole created by the removal of glycans 197, 276, 363, and a glycan from the 460-463 region. Therefore, the computational prediction of glycan processing would be useful to aid laboratory investigations, and this preliminary investigation shows promise as one method to do so.

5. Conclusion

We shall not cease from exploration, and the end of all our exploring will be to arrive where we started and know the place for the first time ~ T S Eliot

This computational study set out to examine the influences of different glycoforms, specifically Man9, Man5 and Complex-glycans, on the accessibility and stability of the HIV-1 Env glycan shield, and, further, to offer a method of predicting the type of glycoform by estimating the degree of glycan-processing based on glycan location and accessibility.

One significant finding to emerge from this study is that complex glycans appear to play an important role in stabilising the Env trimer. This is notable, because trimer stability is important for eliciting neutralising antibodies. Furthermore, the Complex-glycans in this study formed substantial interprotomer hydrogen bonding interactions at the trimer apex, which appear to contribute to preventing the scissoring phenomenon that was observed on the Man9-Trimer, which, comparatively, did not have substantial apex interprotomer interactions. This may be an important consideration for vaccine and immunogen design if computational methods are used since dynamical differences were observed here and, thus, a more representative trimer (such as the Mixed-Trimer) may be more suited to this task.

Interestingly, the Complex-glycans had a much greater impact on the dynamics of the glycan shield, compared to the Man9 glycans, than the Man5 glycans did; in fact, the Man5 and Man9 glycans showed no significant differences in their hydrogen bonding involvement across frames. An implication of this is that previous MD studies using uniformly high mannose glycan forms are likely to agree in terms of their hydrogen bonding dynamics. However, the immense impact of the Complex-glycans on the dynamics of the glycan shield highlights a weakness in these prior studies. It may be useful to repeat some of these studies using a glycan shield comprising a representative mix of glycans. Notwithstanding the relative simplicity of the Mixed-Trimer model in this study, this model gives a better insight into the dynamics of a more natively glycosylated trimer than any previous model.

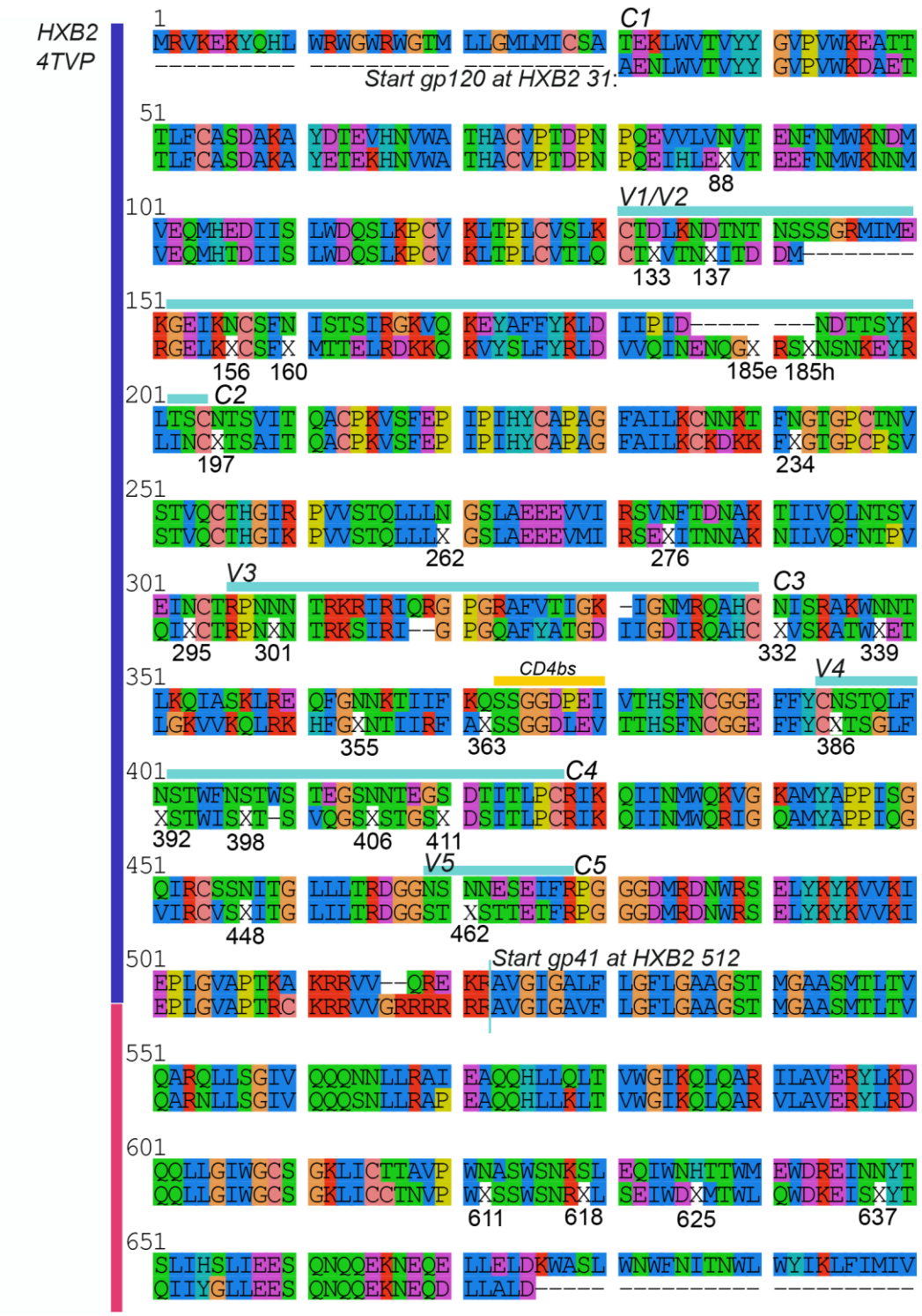
Given the impact of the bi-antennary Complex-glycans used in this study, the question of how hybrid glycans, and tri- and tetra-antennary complex glycans might affect the glycan shield dynamics arises. A natural progression of this work would be to further diversify the broad classifications of glycans used to define the Mixed-Trimer. Further, it may be interesting to compare the Env glycan shields from different cell types, which would vary in their complex glycan forms (Pritchard, Harvey, *et al.*,

2015), in order to understand whether some cells may allow better neutralisation of HIV-1. Additionally, comparing the dynamics of different strains, or recombinant with virion-derived trimers as in Cao *et al*. (2018), as well as isolates from early and late infection, would be useful in expanding our knowledge of how the dynamics of the HIV-1 Env glycan shield varies.

This study also found that the EASA may be useful as a rudimentary tool for predicting whether a glycan would be high mannose or complex; additionally, the EASA appears to be more effective than the SASA as a proxy for predicting the level of glycan processing. This finding may prove useful in future studies, since a computational tool for prediction is much less costly and time-consuming than a mass spectrometry experiment.

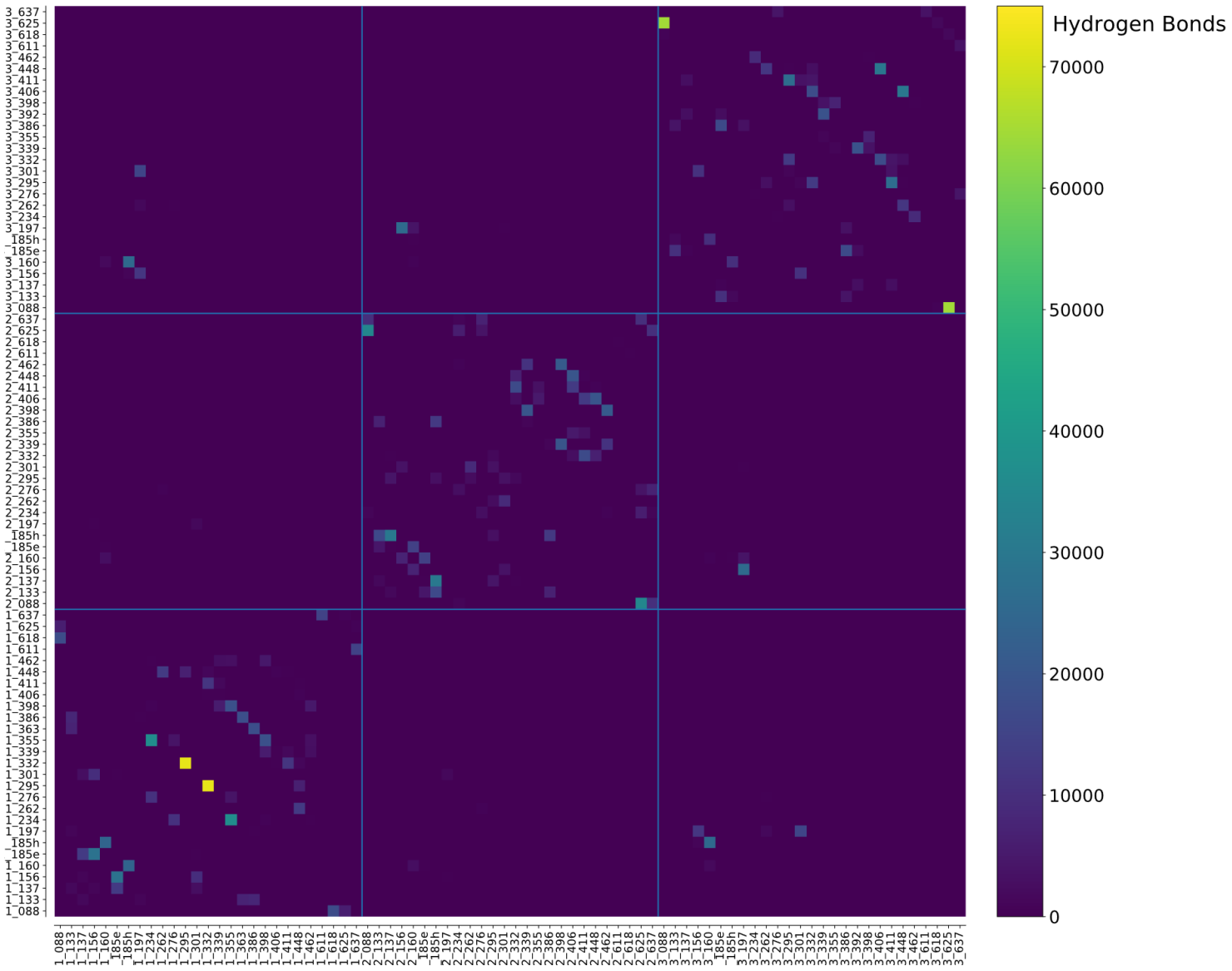
Although a relatively simple mixed-glycan model was used, and the length of a molecular dynamics study can always be extended, this thesis has provided a deeper insight into the HIV-1 Env glycan shield, offering a new perspective of its dynamics. By comparing a uniformly glycosylated trimer with a more representative heterogeneously glycosylated trimer, it became apparent that a more representative glycan shield has different dynamics to a uniform glycan shield. Additionally, the role of physical accessibility in glycan processing was highlighted, and a preliminary method based on the enzyme accessibility was explored as a glycan-processing prediction method. These results provide a useful foundation for future structural and molecular dynamics work to build upon. Overall, an increased understanding of the dynamics of the HIV-1 Env trimer, in relation to its adaptations, mutations, and neutralisation by antibodies, will be helpful for designing immunogens that may eventually lead to an HIV-1 vaccine.

Appendix



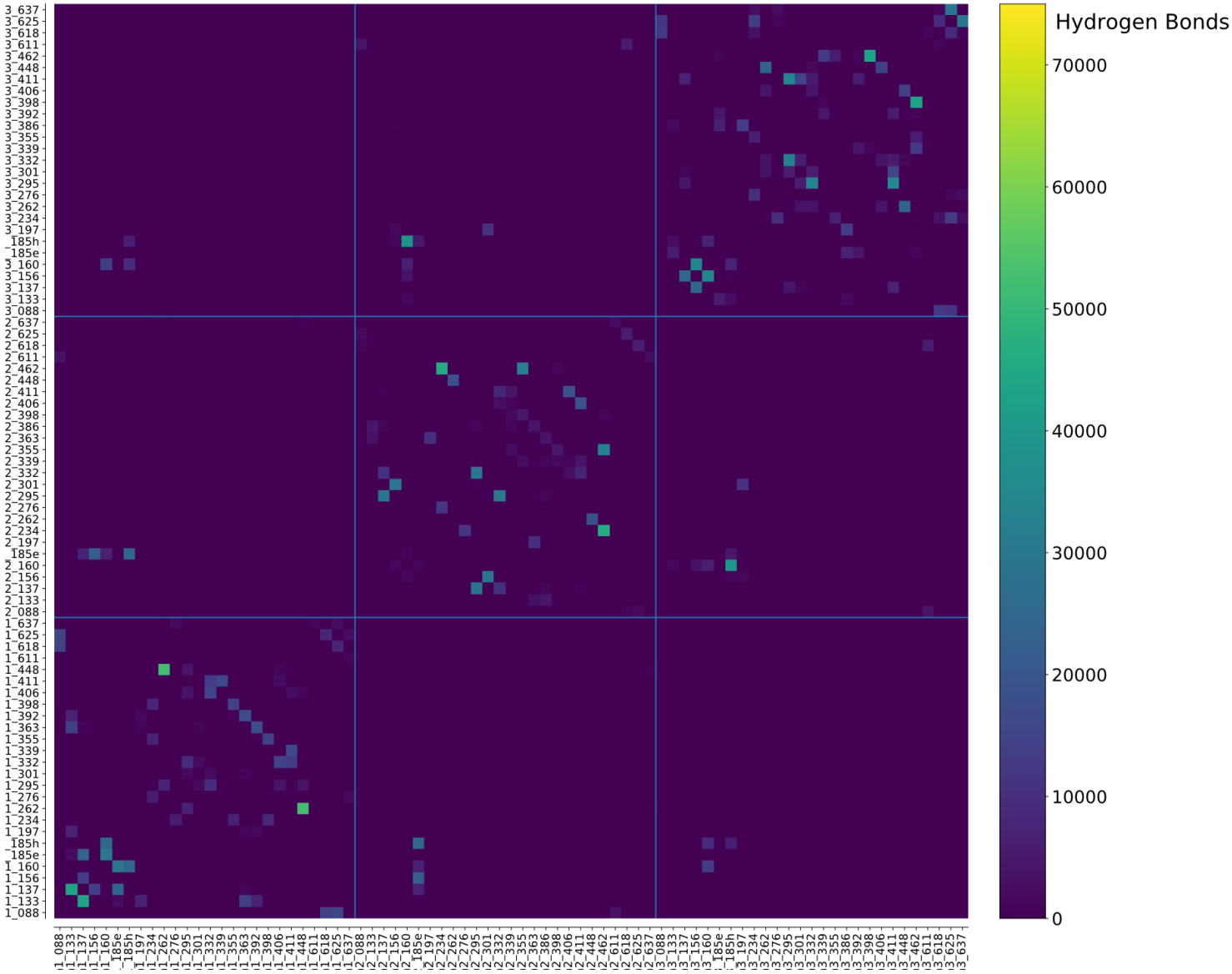
Supplementary Figure 1: HXB2 aligned 4TVP template with PNGSs highlighted

The 4TVP template sequence was aligned to the HXB2 sequence (downloaded from https://www.hiv.lanl.gov/content/index) in SeaView (Gouy, Guindon and Gascuel, 2010). Annotations were added to show HIV-1 regions, and numbers of PNGSs. Although shown as X in the alignment, since they are modified from the standard amino acid code by the software when a glycan is added, PNGSs take on the form **N**XT/S (where X is any amino acid except for proline). Blue and pink lines on the left indicate the gp120 (blue) and gp41 (pink) subparts.



Supplementary Figure 2: Heatmap showing glycan-glycan hydrogen bonding on Man9-Trimer

The number of hydrogen bonds between glycans on the Man9-Trimer were plotted onto a heatmap in order to investigate the distribution of hydrogen bonds across individual glycans, with a grid used to represent protomer divisions. Main upward-right diagonal shows intra-protomer interactions whereas the remaining blocks depict interprotomer interactions. The interactions between glycans 295 and 332 of protomer 1, shown in yellow, contribute the highest number of hydrogen bonds between any two glycans on the Man9-Trimer. Note: glycans 392 of Protomer 1, 363 and 392 of Protomer 2, and 363 of Protomer 3 were not added to this trimer (See Table 3), and thus are not shown on the heatmap.



Supplementary Figure 3: Heatmap showing glycan-glycan hydrogen bonding on Mixed-Trimer

The number of hydrogen bonds between glycans on the Mixed-Trimer were plotted onto a heatmap in order to investigate the distribution of hydrogen bonds across individual glycans, with a grid used to represent protomer divisions. Main upward-right diagonal shows intra-protomer interactions whereas the remaining blocks depict interprotomer interactions. Note: glycans 386 (Man9) and 462 (Complex) of Protomer 1, 185h (Complex) and 392 (Man9) of Protomer 2, and 363 (Man9) of Protomer 3 were not added to this trimer (See Table 3), and thus are not shown on the heatmap.

Supplementary Table 1: Top 10 Ranked RMSF values used for delineating outliers

RMSF values were ranked highest to lowest along with their sequence numbers. The sequence numbers which corresponded to the outliers caused by lack of positional restraints were excluded from the Mann-Whitney U test (italicised and highlighted in dark grey). The highest non-outlying value was 6.6927; thus 6.7 was set as the cutoff, and only RMSF values below this were used in statistical analysis.

Man9-Trimer Sequence number	**RMSF**	**Mixed-Trimer Sequence number**	**RMSF**
1 750	*13.53*	*1 750*	*9.26*
1 751	*12.92*	*1 751*	*8.80*
1 752	*11.70*	*1 752*	*8.30*
1 753	*10.66*	*1 753*	*7.85*
1 754	*9.35*	*1 754*	*6.64*
482	*9.07*	*1 749*	*5.87*
483	*8.28*	372	5.72
1 755	*8.26*	*1 755*	*5.36*
484	*6.94*	373	5.35
1 115	6.69	481	5.23

Supplementary Table 2: Ranking of glycan percentage contribution to hydrogen bonding

Man9-Trimer			Mixed-Trimer		
Glycan	**Number of hydrogen bonds**	**% contribution to hydrogen bonding**	**Glycan**	**Number of hydrogen bonds**	**% contribution to hydrogen bonding**
1_332_m9	85 525	4.01	1_137_c	91 363	3.43
1_295_m9	79 020	3.70	2_462_c	79 922	3.00
3_625_m9	66 264	3.10	3_295_m9	79 294	2.98
3_88_m9	65 493	3.07	1_160_m9	77 144	2.89
1_355_m9	62 174	2.91	1_133_m9	75 259	2.82
2_185h_m9	61 914	2.90	3_160_m9	71 527	2.68
2_625_m9	54 704	2.56	3_411_m9	69 310	2.60
1_185h_m9	51 251	2.40	2_185e_c	68 089	2.56
1_234_m9	48 675	2.28	1_185h_c	66 837	2.51
3_406_m9	47 252	2.21	3_462_c	64 586	2.42
2_88_m9	46 171	2.16	3_625_m9	64 266	2.41
3_448_m9	44 781	2.10	3_156_m9	63 896	2.40
3_295_m9	43 596	2.04	3_185h_c	62 637	2.35
1_185e_m9	41 543	1.95	2_295_m9	61 025	2.29
2_398_m9	40 942	1.92	1_262_m9	60 175	2.26
2_406_m9	40 593	1.90	1_448_m9	58 078	2.18
3_160_m9	39 596	1.86	2_234_m9	58 065	2.18
1_156_m9	39 556	1.85	1_185e_c	56 791	2.13
3_411_m9	37 640	1.76	3_332_m9	54 705	2.05
2_156_m9	37 468	1.76	2_160_m9	54 419	2.04
3_332_m9	37 375	1.75	2_332_m9	51 033	1.92
2_137_m9	36 724	1.72	3_398_m9	44 718	1.68
2_411_m9	33 948	1.59	2_137_c	42 163	1.58
3_197_m9	33 922	1.59	2_301_m9	40 772	1.53
2_462_m9	32 953	1.54	2_355_m5	40 246	1.51
2_339_m9	31 983	1.50	1_332_m9	40 090	1.50
3_301_m9	31 174	1.46	3_448_m9	40 010	1.50
1_197_m9	30 539	1.43	3_137_c	38 235	1.43
3_185e_m9	30 527	1.43	1_156_m9	37 231	1.40
2_160_m9	30 343	1.42	3_234_m9	36 628	1.37
1_398_m9	29 520	1.38	1_363_m9	35 904	1.35
2_133_m9	29 513	1.38	3_637_m5	35 902	1.35
1_160_m9	29 401	1.38	3_262_m9	35 494	1.33
2_332_m9	28 292	1.33	2_156_m9	34 309	1.29
2_637_m9	28 230	1.32	1_88_m9	33 690	1.26
2_448_m9	27 239	1.28	3_301_m9	32 860	1.23
1_386_m9	26 409	1.24	1_411_m9	32 279	1.21
1_363_m9	25 482	1.19	2_411_m9	31 873	1.20
3_392_m9	24 714	1.16	1_392_m9	28 124	1.06

3_386_m9	24 067	1.13	1_295_m9	27 449	1.03
3_156_m9	23 782	1.11	3_618_c	27 048	1.02
1_88_m9	23 750	1.11	1_625_m9	26 052	0.98
3_339_m9	23 521	1.10	3_197_m5	25 183	0.95
1_448_m9	21 601	1.01	3_88_m9	24 393	0.92
2_386_m9	20 829	0.98	2_406_m9	23 387	0.88
2_185e_m9	19 733	0.92	1_355_m5	22 879	0.86
1_133_m9	18 464	0.87	1_398_m9	22 680	0.85
1_137_m9	18 194	0.85	1_406_m9	22 515	0.84
1_618_m9	17 587	0.82	1_618_c	22 495	0.84
3_262_m9	17 553	0.82	3_386_m9	22 280	0.84
3_133_m9	17 010	0.80	3_406_m9	22 120	0.83
2_301_m9	16 139	0.76	1_234_m9	21 910	0.82
1_637_m9	15 690	0.74	2_448_m9	19 072	0.72
1_276_m9	15 485	0.73	2_262_m9	18 801	0.71
1_611_m9	14 951	0.70	3_339_m9	18 677	0.70
2_295_m9	14 815	0.69	3_185e_c	18 545	0.70
2_276_m9	14 645	0.69	2_363_m9	17 615	0.66
1_301_m9	14 619	0.68	1_339_m9	16 933	0.64
1_411_m9	13 962	0.65	3_392_m9	16 549	0.62
1_262_m9	13 533	0.63	2_618_c	13 682	0.51
3_185h_m9	12 079	0.57	3_276_m5	13 043	0.49
2_234_m9	11 618	0.54	3_355_m5	12 852	0.48
2_262_m9	11 495	0.54	2_386_m9	12 520	0.47
3_398_m9	11 367	0.53	2_276_m5	12 122	0.45
1_462_m9	11 059	0.52	3_133_m9	11 763	0.44
1_339_m9	10 778	0.50	2_339_m9	10 970	0.41
3_462_m9	9 568	0.45	3_611_c	10 905	0.41
3_234_m9	9 260	0.43	2_197_m5	10 113	0.38
2_355_m9	8 889	0.42	1_197_m5	10 107	0.38
3_355_m9	8 195	0.38	2_625_m9	8 304	0.31
3_637_m9	7 916	0.37	2_133_m9	8 293	0.31
1_625_m9	7 194	0.34	1_276_m5	7 955	0.30
3_137_m9	6 632	0.31	2_398_m9	7 413	0.28
3_276_m9	4 783	0.22	2_88_m9	6 817	0.26
3_611_m9	3 661	0.17	2_611_c	6 338	0.24
3_618_m9	2 662	0.12	1_637_m5	5 420	0.20
2_197_m9	2 226	0.10	1_301_m9	5 234	0.20
1_406_m9	1 789	0.08	2_637_m5	4 003	0.15
2_618_m9	498	0.02	1_611_c	1 426	0.05
2_611_m9	457	0.02			

Supplementary Methods: Algorithms written for this project

A. Core algorithm used for parsing the CPPTRAJ hydrogen bonding output

Note to the reader: If you are not familiar with code, the lines beginning with # give a general description of the steps in the algorithm below.

This code is related to Section 3.4.1. Quantifying the glycan interactions from simulations.

```
def hbond_csv(hbonds, dct, rev):
    '''
    INPUT:
    - hbonds: converted csv output from the hbonds module in CPPTRAJ
    - dct: glycan dictionary
    - rev: & reverse dictionary mapping residues to glycans

    OUPUT
    - A csv file for every glycan, which contains a column for every interaction
      that glycan has with another glycan
        --> filename = glycan eg: N88
        --> header names = glycan interactions for that glycan eg: N88-N133
        --> each row a frame of summed hydrogen bonds in that glyc-glyc intxn
    '''
    hbond_table = pd.read_csv(hbonds, header=0)
    colheaders = list(hbond_table)
    summed_dfs = {}
    # {N88:{N88-N133:sdfs}, {N88-N37:sdfs}, N133 etc}

    for glycan, res in dct.items():
        # eg dct: {'m1_88': ['482', '483', '484', ..., '491', '492'], ...}
        #print(glycan)
        summed_dfs[glycan] = {}
        # creating a dictionary within a dictionary, populated by each pass of
        # the for loop
        interglyc = []  # header names of interglycan hydrogen bonding
        # --> to use as a filter with the giant table
        # eg: ["4YB_2607@O3-OMA_2799@O4-H4O", ...]
        glyc_intxn_heads = []  # glycan names for the interglycan interactions
        # --> to make human sense of each interaction and relate it back to the
        # glycans we are dealing with
        # eg: ["m1_133-m1_137", "m1_133-m2_88"]
        for resid in res:
            # print(resid)
            for colname in colheaders:
                # eg colname: 4YB_48@O3-4YB_26@N2-H2N
                # eg extract_resids output: [48, 26]
                print(resid, colname)
                if resid in colname:
                    print("resid {} in: {}".format(resid, colname))
                    res1, res2 = (extract_resids(colname))  # res1 & 2 are ints
                    print("res1: ", res1)
                    print("res2: ", res2)

                    if resid != str(res1) and resid != str(res2):
                        pass

                    else:
```

```
                    if resid == str(res1):
                        print("resid == res1")
                        glycan_1 = glycan
                        glycan_2 = rev[str(res2)]
                        print(glycan_1, glycan_2)

                    if resid == str(res2):
                        print("resid == res2")
                        glycan_1 = glycan
                        glycan_2 = rev[str(res1)]
                        print(glycan_1, glycan_2)

                    if glycan_1 == glycan_2: # filter out m1_133 == m1_133 etc
                        print("not including as intra-glycan interaction")
                        # exclude column if glycans are the same through
                        # if/else statement
                    else:
                        # we want to keep it
                        if colname in interglyc:
                          # check if it's already there
                            pass
                        else:
                            # if not, add it to the interglyc list in form N88-N133
                            interglyc.append(colname)
                            glyc_intxn_heads.append(str(glycan_1) + "-" + str(glycan_2))

            else:
                pass

    if interglyc != []:

        df_interglyc = hbond_table[interglyc]
        # This is filtered down to just the interglycan hydrogen bonds
        # So then need to see where the glycans are the same (eg: N88 & N133)
        # and add these columns together

        # Get the indices for every unique glycan interaction into a dictionary
        intxn_indices = defaultdict(list)
        for index, intxn in enumerate(glyc_intxn_heads):
            intxn_indices[intxn].append(index)

        # Use indices from above for each unique hbond to look up these columns
        # and sum them (eg summing all N88-N133 interactions)
        # --> summing by frame for each glyc-glyc interaction
        for intxn in intxn_indices:
            print("-->", intxn)
            print(type(intxn))
            sum_df = df_interglyc.iloc[:, intxn_indices[intxn]].sum(axis=1)
            summed_dfs[glycan][intxn] = sum_df
            print(sum_df)
        # Join the summed dfs into one df per glycan to show all its intxns
        joined_dfs = pd.concat(summed_dfs[glycan], axis=1)

        with open("{}.csv".format(glycan), 'w', newline='') as csvfile:
            writer = csv.DictWriter(csvfile, fieldnames=False)
            joined_dfs.to_csv(csvfile, index=False)

print("done :) \nCheck working directory for output csv files")
```

B. Core algorithms for calculating the hydrogen bonding by glycan

This code is related to Section 3.4.1. Quantifying the glycan interactions from simulations.

Step 1: Command Line

- `for file in *.csv; do python add.py -hb ${file} >> mixed_sums.txt; done #used on the .csv output by glycan from A`

```
Main contents of add.py:
    hbond_table = pd.read_csv(sum_hbonds, header=0)
    print(hbond_table.sum())
```

- `grep -v dtype mixed_sums.csv > mix_clean.csv`

Step 2: Python

```
man9_df = pd.read_csv('man9_clean.csv', header = None)
mixed_df = pd.read_csv('mix_clean.csv', header = None)

#making them sets in order to use operators on them for set theory
man9_set = OrderedSet(man9_df[0])
mixed_set = OrderedSet(mixed_df[0])

common_interactions = man9_set & mixed_set
man9_only_interactions = man9_set - mixed_set
mixed_only_interactions = mixed_set - man9_set

man9_dict = dict(zip(list(man9_df[0]), list(man9_df[1])))
mixed_dict = dict(zip(list(mixed_df[0]), list(mixed_df[1])))

# Make a dictionary of the common interactions in the form:
# {'m1_262-m1_295': [100, 7752], 'm1_625-m1_637': [672, 2401]
# Then change to a dataframe for checking & collating
commonintxns_dict = collections.defaultdict(list)
for interaction in common_interactions:
    commonintxns_dict[interaction].append(man9_dict[interaction])
    commonintxns_dict[interaction].append(mixed_dict[interaction])
commonintxns_df = pd.DataFrame.from_dict(commonintxns_dict, orient='index')

# Make Man9 & mixed dictionaries of the form:
# {'m1_262-m1_295': [100], ...}
# Then change to a dataframes for checking & collating

man9onlyintxns_dict = collections.defaultdict(list)
for interaction in man9_only_interactions:
    man9onlyintxns_dict[interaction].append(man9_dict[interaction])
m9onlydf = pd.DataFrame.from_dict(man9onlyintxns_dict, orient='index')

mixedonlyintxns_dict = collections.defaultdict(list)
for interaction in mixed_only_interactions:
    mixedonlyintxns_dict[interaction].append(mixed_dict[interaction])
mixedonlydf = pd.DataFrame.from_dict(mixedonlyintxns_dict, orient='index')

#Here: do check of glycan pairs and sums

#First reindex dfs
commonintxns_df_reindexed = commonintxns_df.reindex(reorder_intxn(commonintxns_df))
```

```
m9onlydf_reindexed = m9onlydf.reindex(reorder_intxn(m9onlydf))
mixedonlydf_reindexed = mixedonlydf.reindex(reorder_intxn(mixedonlydf))

#Then reorder by reindex so that pairs end up next to each other
paired_commonintxndf = commonintxns_df_reindexed.sort_index(axis=0)
paired_m9intxndf = m9onlydf_reindexed.sort_index(axis=0)
paired_mixintxndf = mixedonlydf_reindexed.sort_index(axis=0)

#Check that each glycan interaction has a partner, and that the number of hbonds is the
same
if check_clause(paired_commonintxndf, 2) == True:
    if check_clause(paired_m9intxndf, 1) == True:
        if check_clause(paired_mixintxndf, 1) == True:

            #remove zeros from 1 column dfs
            paired_m9intxndf = zero_count(paired_m9intxndf, 0)[0]
            paired_mixintxndf = zero_count(paired_mixintxndf, 0)[0]

            # remove zeros in m9 column of common intxns
            paired_commonintxndf, add_to_mixed = zero_count(paired_commonintxndf, 0)[:2]
            # remove zeros in mixed column
            paired_commonintxndf, add_to_m9 = zero_count(paired_commonintxndf, 1)[:2]


            #add in the values column to the relevant 1 column df, eg:
#                       add_to_m9 [                   0  1
#             m1_332-m1_448   1185  0
#             m1_332-m1_448   1185  0
#             m2_137-m2_185e   391  0
#             m2_137-m2_185e   391  0
#             m3_386-m3_462     18  0
#             m3_386-m3_462     18  0]

            add_to_m9_df = pd.Series.to_frame(add_to_m9[0][0])
            paired_m9intxndf = paired_m9intxndf.append(add_to_m9_df)
            add_to_mixed[0][0] = add_to_mixed[0][1]
           # Little bit roundabout, but needed to join
            # to paired_mixintxndf on the same column index
            add_to_mixed_df = pd.Series.to_frame(add_to_mixed[0][0])
            paired_mixintxndf = paired_mixintxndf.append(add_to_mixed_df)

            # Take every second row of the dataset, so that you rule out duplicates
             #(since paired)
            clean_m9_df = paired_m9intxndf.iloc[::2]
            clean_mixed_df = paired_mixintxndf.iloc[::2]
            clean_common_df = paired_commonintxndf.iloc[::2]
            #Check:
            print("\ncommon sum \nm9: ", clean_common_df.sum()[0], "\nmixed: ",
             clean_common_df.sum()[1])
            print("\nm9 only sum: ", clean_m9_df.sum()[0])
            print("mixed only sum: ", clean_mixed_df.sum()[0])

            #Add up the number of hydrogen bonds for each glycan:
            m9_hbond_by_glyc_df =(glycan_hbond_count(clean_m9_df, 0,
             M9_GLYCANS)).add(glycan_hbond_count(clean_common_df, 0,
             M9_GLYCANS)).transpose().sort_values(by=[0], ascending=False)
            mix_hbond_by_glyc_df = glycan_hbond_count(clean_mixed_df, 0,
             MIX_GLYCANS).add(glycan_hbond_count(clean_common_df, 1,
             MIX_GLYCANS)).transpose().sort_values(by=[0], ascending=False)
```

C. Core algorithm for sorting glycan hydrogen bonds by region

This code is related to Section 3.4.2. Assessing how glycan-glycan hydrogen bonds cluster into the defined regions of HIV-1 Env.

```
#Regions dictionary (for glycans)
regions_dict = {"C1": ["088"], "V1/V2": ["133", "137", "156", "160", "185e", "185h"],
"C2": ["197", "234", "262", "276", "295"], "V3": ["301"], "C3": ["332", "339", "355",
"363"], "V4": ["386", "392", "398", "406", "411"], "C4": ["448"], "V5": ["462"],
"gp41": ["611", "618", "625", "637"]}
res_regions_dict = reverso(regions_dict)

hbond_table = pd.read_csv(hbonds_table, header = 0)
hbond_regions = hbond_table

reg1 = []
reg2 = []

for glycan in hbond_regions["Glycan1"]:
    a = (glycan.split("_")[0].strip("m"), res_regions_dict[glycan.split("_")[1]])
    reg1.append("_".join(a))

for glycan in hbond_regions["Glycan2"]:
    b = (glycan.split("_")[0].strip("m"), res_regions_dict[glycan.split("_")[1]])
    reg2.append("_".join(b))

hbond_regions["Glycan1"] = reg1
hbond_regions["Glycan2"] = reg2
```

D. Core algorithm for percentage hydrogen bond presence over the simulation

This code is related to Section 3.4.4. Investigating hydrogen bond longevity versus strength of bonding.

```
#Similar to B.1. this script is used with a batch script.
# Used on the csv outputs from B.2.
int_table = pd.read_csv(hbond_by_frame, header=0)
    # print(int_table)
    dict_hbond_counts = collections.defaultdict(list)

    for header in int_table:
        no_hbonds_for_intxn = 0
        for x in int_table[header]:
            if x >= 1:
                no_hbonds_for_intxn +=1
        dict_hbond_counts[header].append(no_hbonds_for_intxn/len(int_table)*100)

    print(pd.DataFrame.from_dict(dict_hbond_counts, orient='index'))
```

E. Core algorithms for percentage glycan involvement in hydrogen bonding over the simulation

This code is related to Section 3.4.4. Investigating hydrogen bond longevity versus strength of bonding.

- `for file in *.csv; do sort -n $file | uniq -c >> frames_present.txt;done`
- `grep -e m1 -e m2 -e m3 frames_present.txt | awk '{print $2}' > list_of_intxns`

- `grep -e m1 -e m2 -e m3 -B 1 frames_present.txt | awk '{print $1}' | grep -w -v 1 | grep -w -v '[^a-zA-Z0-9]' > frames_present_ordered_nos`
- `awk -F- '{print $1}' list_of_intxns > glycs_ordered`
- `paste -d, glycs_ordered frames_present_ordered_nos > no_frames_not_glyc.csv`
- `#NB: Since this algorithm finds the frames where the glycan is not involved, a further step is then necessary`

F. Core algorithm for the glycan processing prediction analysis

This code is related to Section 3.6. Investigating Enzyme and Solvent Accessible Surface Area as a Proxy for Glycan Processing

```
m9_sasa_df = pd.read_csv("0MA_SASAs.csv", header=None)
print(m9_sasa_df)
m9_sasa_df = m9_sasa_df.T
print(m9_sasa_df.iloc[0:3])

sasa_df = pd.DataFrame()

acc_glyc = 0
for x in m9_glycans:
    print(x)
    print(m9_sasa_df.loc[:, acc_glyc])
    glyc_frames = [m9_sasa_df.iloc[:, acc_glyc], m9_sasa_df.iloc[:, acc_glyc+1],
m9_sasa_df.iloc[:, acc_glyc+2]]
    glyc_values = pd.concat(glyc_frames)
    print(glyc_values)
    frames = [sasa_df, glyc_values]
    sasa_df = pd.concat(frames, axis=1)
    acc_glyc += 3

m9_glycans = ['1_88_m9', '1_133_m9', '1_137_c', '1_156_m9', '1_160_m9', '1_185e_c',
'1_185h_c', '1_197_m5', '1_234_m9', '1_262_m9', '1_276_m5', '1_295_m9', '1_301_m9',
'1_332_m9', '1_339_m9', '1_355_m5', '1_363_m9', '1_386_m9', '1_398_m9', '1_406_m9',
'1_411_m9', '1_448_m9', '1_462_c', '1_611_c', '1_618_c', '1_625_m9', '1_637_m5',
'2_88_m9', '2_133_m9', '2_137_c', '2_156_m9', '2_160_m9', '2_185e_c', '2_185h_c',
'2_197_m5', '2_234_m9', '2_262_m9', '2_276_m5', '2_295_m9', '2_301_m9', '2_332_m9',
'2_339_m9', '2_355_m5', '2_386_m9', '2_398_m9', '2_406_m9', '2_411_m9', '2_448_m9',
'2_462_c', '2_611_c', '2_618_c', '2_625_m9', '2_637_m5', '3_88_m9', '3_133_m9',
'3_137_c', '3_156_m9', '3_160_m9', '3_185e_c', '3_185h_c', '3_197_m5', '3_234_m9',
'3_262_m9', '3_276_m5', '3_295_m9', '3_301_m9', '3_332_m9', '3_339_m9', '3_355_m5',
'3_386_m9', '3_392_m9', '3_398_m9', '3_406_m9', '3_411_m9', '3_448_m9', '3_462_c',
'3_611_c', '3_618_c', '3_625_m9', '3_637_m5']

sasa_df.columns = m9_glycans

C_frames = []
m5_frames = []
m9_frames = []
for col in sasa_df.columns:

    if glycoform_dict[col.split("_")[1]] == "C":
        C_frames.append(col)
    if glycoform_dict[col.split("_")[1]] == "M5":
        m5_frames.append(col)
    if glycoform_dict[col.split("_")[1]] == "M9":
        m9_frames.append(col)
```

References

Atun, R. *et al.* (2016) 'Long-term financing needs for HIV control in sub-Saharan Africa in 2015-2050: a modelling study.', *BMJ open*, 6(3), p. e009656. doi: 10.1136/bmjopen-2015-009656.

Barré-Sinoussi, F. *et al.* (1983) 'Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).', *Science (New York, N.Y.)*, 220(4599), pp. 868–71.

Behrens, A.-J. *et al.* (2016) 'Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein', *Cell Reports*, 14, pp. 2695–2706. doi: 10.1016/j.celrep.2016.02.058.

Behrens, A.-J. *et al.* (2018) 'Integrity of Glycosylation Processing of a Glycan-Depleted Trimeric HIV-1 Immunogen Targeting Key B-Cell Lineages', *Journal of Proteome Research*, 17(3), pp. 987–999. doi: 10.1021/acs.jproteome.7b00639.

Binley, J. M. *et al.* (2000) 'A Recombinant Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Complex Stabilized by an Intermolecular Disulfide Bond between the gp120 and gp41 Subunits Is an Antigenic Mimic of the Trimeric Virion-Associated Structure', *Journal of Virology*, 74(2), pp. 627–643. doi: 10.1128/JVI.74.2.627-643.2000.

Brandenberg, O. F. *et al.* (2015) 'The HIV-1 Entry Process: A Stoichiometric View', *Trends in Microbiology*, 23(12), pp. 763–774. doi: 10.1016/j.tim.2015.09.003.

Brenchley, J. M. *et al.* (2004) 'CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract.', *The Journal of experimental medicine*. Rockefeller University Press, 200(6), pp. 749–59. doi: 10.1084/jem.20040874.

Briggs, J. A. G. *et al.* (2003) 'Structural organization of authentic, mature HIV-1 virions and cores.', *The EMBO journal*. European Molecular Biology Organization, 22(7), pp. 1707–15. doi: 10.1093/emboj/cdg143.

Buonaguro, L., Tornesello, M. L. and Buonaguro, F. M. (2007) 'Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications.', *Journal of virology*. American Society for Microbiology (ASM), 81(19), pp. 10209–19. doi: 10.1128/JVI.00872-07.

Cao, L. *et al.* (2017) 'Global site-specific N-glycosylation analysis of HIV envelope glycoprotein', *Nature Communications*. Nature Publishing Group, 8, pp. 1–13. doi: 10.1038/ncomms14954.

Cao, L. *et al.* (2018) 'Differential processing of HIV envelope glycans on the virus and soluble recombinant trimer', *Nature Communications*. Springer US, 9(1), p. 3693. doi: 10.1038/s41467-018-06121-4.

Case, D. A. *et al.* (2014) 'AMBER 2014', *University of California, San Francisco*. University of California, San Francisco.

Chan, D. C. and Kim, P. S. (1998) 'HIV entry and its inhibition.', *Cell*. Elsevier, 93(5), pp. 681–4. doi:

10.1016/S0092-8674(00)81430-0.

Chang, V. T. *et al.* (2007) 'Glycoprotein structural genomics: solving the glycosylation problem.', *Structure (London, England : 1993)*. Elsevier, 15(3), pp. 267–73. doi: 10.1016/j.str.2007.01.011.

Checkley, M. A., Luttge, B. G. and Freed, E. O. (2011) 'HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation', *Journal of Molecular Biology*, 410(4), pp. 582–608. doi: 10.1016/j.jmb.2011.04.042.

Coss, K. P. *et al.* (2016) 'HIV-1 Glycan Density Drives the Persistence of the Mannose Patch within an Infected Individual.', *Journal of virology*. American Society for Microbiology, 90(24), pp. 11132–11144. doi: 10.1128/JVI.01542-16.

Crispin, M., Ward, A. B. and Wilson, I. A. (2018) 'Structure and Immune Recognition of the HIV Glycan Shield', *Annual Review of Biophysics*, 47(1), pp. 499–523. doi: 10.1146/annurev-biophys-060414-034156.

Crooks, E. T. *et al.* (2015) 'Vaccine-Elicited Tier 2 HIV-1 Neutralizing Antibodies Bind to Quaternary Epitopes Involving Glycan-Deficient Patches Proximal to the CD4 Binding Site', *PLOS Pathogens*. Edited by A. Trkola. Public Library of Science, 11(5), p. e1004932. doi: 10.1371/journal.ppat.1004932.

Curlin, M. E. *et al.* (2010) 'HIV-1 envelope subregion length variation during disease progression.', *PLoS pathogens*. Public Library of Science, 6(12), p. e1001228. doi: 10.1371/journal.ppat.1001228.

Domingo, E. *et al.* (1978) 'Nucleotide sequence heterogeneity of an RNA phage population', *Cell*. Cell Press, 13(4), pp. 735–744. doi: 10.1016/0092-8674(78)90223-4.

Doores, K. J. *et al.* (2010) 'Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(31), pp. 13800–5. doi: 10.1073/pnas.1006498107.

Doores, K. J. (2015) 'The HIV glycan shield as a target for broadly neutralizing antibodies.', *The FEBS journal*, 282(24), pp. 4679–91. doi: 10.1111/febs.13530.

Duan, H. *et al.* (2018) 'Glycan Masking Focuses Immune Responses to the HIV-1 CD4-Binding Site and Enhances Elicitation of VRC01-Class Precursor Antibodies', *Immunity*. Elsevier Inc., pp. 1–11. doi: 10.1016/j.immuni.2018.07.005.

Durrant, J. D. and McCammon, J. A. (2011) 'Molecular dynamics simulations and drug discovery.', *BMC biology*, 9, p. 71. doi: 10.1186/1741-7007-9-71.

Earl, P. L., Doms, R. W. and Moss, B. (1990) 'Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 87, pp. 648–52.

Eigen, M. and Schuster, P. (1977) ‘A principle of natural self-organization - Part A: Emergence of the hypercycle’, *Naturwissenschaften*. Springer-Verlag, 64(11), pp. 541–565. doi: 10.1007/BF00450633.

Engelman, A. and Cherepanov, P. (2012) ‘The structural biology of HIV-1: mechanistic and therapeutic insights.’, *Nature reviews. Microbiology*. NIH Public Access, 10(4), pp. 279–90. doi: 10.1038/nrmicro2747.

Feng, Y. *et al.* (2016) ‘Thermostability of Well-Ordered HIV Spikes Correlates with the Elicitation of Autologous Tier 2 Neutralizing Antibodies’, *PLOS Pathogens*, 12(8), p. e1005767.

Ferreira, R. *et al.* (2018) ‘Structural Rearrangements Maintain the Glycan Shield of an HIV-1 Envelope Trimer After the Loss of a Glycan’, *Scientific Reports*. Springer US, (May), pp. 1–15. doi: 10.1038/s41598-018-33390-2.

Frankel, A. D. and Young, J. A. T. (1998) ‘HIV-1: Fifteen Proteins and an RNA’, *Annual Review of Biochemistry*, 67(1), pp. 1–25. doi: 10.1146/annurev.biochem.67.1.1.

Freed, E. O. (2015) ‘HIV-1 assembly, release and maturation’, *Nature Reviews Microbiology*, 13(8), pp. 484–496. doi: 10.1038/nrmicro3490.

Gallo, R. C. *et al.* (1983) ‘Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS).’, *Science (New York, N.Y.)*, 220(4599), pp. 865–7.

Garces, F. *et al.* (2015) ‘Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans.’, *Immunity*. Elsevier, 43(6), pp. 1053–63. doi: 10.1016/j.immuni.2015.11.007.

Geijtenbeek, T. B. *et al.* (2000) ‘DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells.’, *Cell*. Elsevier, 100(5), pp. 587–97. doi: 10.1016/S0092-8674(00)80694-7.

Gift, S. K. *et al.* (2017) ‘Functional Stability of HIV-1 Envelope Trimer Affects Accessibility to Broadly Neutralizing Antibodies at its Apex.’, *Journal of virology*, (October), p. JVI.01216-17. doi: 10.1128/JVI.01216-17.

Gorman, J. *et al.* (2016) ‘Structures of HIV-1 Env V1V2 with broadly neutralizing antibodies reveal commonalities that enable vaccine design.’, *Nature structural & molecular biology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 23(1), pp. 81–90. doi: 10.1038/nsmb.3144.

Gouy, M., Guindon, S. and Gascuel, O. (2010) ‘SeaView version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building.’, *Molecular Biology and Evolution*, 27(2), pp. 221–224.

Gray, E. S. *et al.* (2011) ‘The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection.’, *Journal of virology*, 85(10), pp. 4828–40. doi: 10.1128/JVI.00198-11.

Gristick, H. B., Wang, H. and Bjorkman, P. J. (2017) 'X-ray and EM structures of a natively glycosylated HIV-1 envelope trimer', *Acta Crystallographica Section D Structural Biology*. International Union of Crystallography, 73(10), pp. 822–828. doi: 10.1107/S2059798317013353.

Grossman, Z. *et al.* (2006) 'Pathogenesis of HIV infection: what the virus spares is as important as what it destroys.', *Nature medicine*. Nature Publishing Group, 12(3), pp. 289–95. doi: 10.1038/nm1380.

Gu, J. *et al.* (2018) 'Molecular dynamics perspective on the thermal stability of mandelate racemase', *Journal of Biomolecular Structure and Dynamics*. Taylor & Francis, 1102, pp. 1–11. doi: 10.1080/07391102.2018.1427631.

Günthard, H. F. and Scherrer, A. U. (2016) 'HIV-1 Subtype C, Tenofovir, and the Relationship With Treatment Failure and Drug Resistance.', *The Journal of infectious diseases*. Oxford University Press, 214(9), pp. 1289–1291. doi: 10.1093/infdis/jiw214.

Guttman, M. *et al.* (2014) 'CD4-Induced Activation in a Soluble HIV-1 Env Trimer', *Structure*, 22(7), pp. 974–984. doi: 10.1016/j.str.2014.05.001.

Huang, C. *et al.* (2005) 'Structure of a V3-containing HIV-1 gp120 core.', *Science (New York, N.Y.)*, 310(5750), pp. 1025–8. doi: 10.1126/science.1118398.

Hubbard, S. J. and Thornton, J. M. (1993) '"NACCESS", Computer Program', *Department of Biochemistry and Molecular Biology, University College London*.

Hunter, J. D. (2007) 'Matplotlib: A 2D Graphics Environment', *Computing in Science & Engineering*, 9(3), pp. 90–95. doi: 10.1109/MCSE.2007.55.

Ihaka, R. and Gentleman, R. (1996) 'R: A Language for Data Analysis and Graphics on JSTOR', *Journal of Computational and Graphical Statistics*, 5(3), pp. 299–314.

John, B. and Sali, A. (2003) 'Comparative protein structure modeling by iterative alignment, model building and model assessment', *Nucleic Acids Research*, 31(14), pp. 3982–3992. doi: 10.1093/nar/gkg460.

Jones, E. *et al.* (2001) *SciPy: Open Source Scientific Tools for Python*. Available at: http://www.scipy.org/.

Julien, J.-P., Lee, J. H., *et al.* (2013) 'Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(11), pp. 4351–6. doi: 10.1073/pnas.1217537110.

Julien, J.-P., Cupo, A., *et al.* (2013) 'Crystal structure of a soluble cleaved HIV-1 envelope trimer.', *Science (New York, N.Y.)*, 342(6165), pp. 1477–83. doi: 10.1126/science.1245625.

Karaveg, K. *et al.* (2005) 'Mechanism of Class 1 (Glycosylhydrolase Family 47) alpha-Mannosidases Involved in N-Glycan Processing and Endoplasmic Reticulum Quality Control', *Journal of Biological Chemistry*,

280(16), pp. 16197–16207. doi: 10.1074/jbc.M505130200.

Keele, B. F. *et al.* (2008) 'Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(21), pp. 7552–7. doi: 10.1073/pnas.0802203105.

Kiln Enterprises Ltd (2016) *Flourish*. Available at: https://flourish.studio/.

Kirschner, K. N. *et al.* (2008) 'GLYCAM06: a generalizable biomolecular force field. Carbohydrates.', *Journal of computational chemistry*. NIH Public Access, 29(4), pp. 622–55. doi: 10.1002/jcc.20820.

Klatzmann, D. *et al.* (1984) 'Selective Tropism of Lymphadenopathy Associated Virus (LAV) for Helper-Inducer T Lymphocytes', *Science*, 225(4657), pp. 59–63.

Klein, F. *et al.* (2013) 'Antibodies in HIV-1 vaccine development and therapy.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 341(6151), pp. 1199–204. doi: 10.1126/science.1241144.

Kong, L., Stanfield, R. L. and Wilson, I. A. (2013) 'Molecular Recognition of HIV Glycans by Antibodies', in *HIV Glycans in Infection and Immunity*, pp. 117–141.

Kornfeld, R. and Kornfeld, S. (1985) 'Assembly of Asparagine-Linked Oligosaccharides', *Annual Review of Biochemistry*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 54(1), pp. 631–664. doi: 10.1146/annurev.bi.54.070185.003215.

Kwong, P. D. *et al.* (1998) 'Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody', *Nature*. Nature Publishing Group, 393(6686), pp. 648–659. doi: 10.1038/31405.

Kwong, P. D. (1999) 'Probability Analysis of Variational Crystallization and Its Application to gp120, The Exterior Envelope Glycoprotein of Type 1 Human Immunodeficiency Virus (HIV-1)', *Journal of Biological Chemistry*, 274(7), pp. 4115–4123. doi: 10.1074/jbc.274.7.4115.

Kwong, P. D., Mascola, J. R. and Nabel, G. J. (2011) 'Rational design of vaccines to elicit broadly neutralizing antibodies to HIV-1.', *Cold Spring Harbor perspectives in medicine*. Cold Spring Harbor Laboratory Press, 1(1), p. a007278. doi: 10.1101/cshperspect.a007278.

Lasky, L. A. *et al.* (1986) 'Neutralization of the AIDS retrovirus by antibodies to a recombinant envelope glycoprotein.', *Science (New York, N.Y.)*, 233(4760), pp. 209–12. doi: 10.1126/science.3014647.

Lau, K. S. *et al.* (2007) 'Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation', *Cell*. Cell Press, 129(1), pp. 123–134. doi: 10.1016/J.CELL.2007.01.049.

Lavine, C. L. *et al.* (2012) 'High-mannose glycan-dependent epitopes are frequently targeted in broad

neutralizing antibody responses during human immunodeficiency virus type 1 infection.', *Journal of virology*. American Society for Microbiology (ASM), 86(4), pp. 2153–64. doi: 10.1128/JVI.06201-11.

Lee, J. H., Ozorowski, G. and Ward, A. B. (2016) 'Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer', *Science*, 351(6277), pp. 1043–1048. doi: doi: 10.1126/science.aad2450.

Lemmin, T. *et al.* (2017) 'Microsecond Dynamics and Network Analysis of the HIV-1 SOSIP Env Trimer Reveal Collective Behavior and Conserved Microdomains of the Glycan Shield', *Structure*. Cell Press, 25(10), pp. 1631-1639.e2. doi: 10.1016/J.STR.2017.07.018.

Leonard, C. K. *et al.* (1990) 'Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells', *Journal of Biological Chemistry*, 265(18), pp. 10373–10382. doi: VL - 265.

Leymarie, N. and Zaia, J. (2012) 'Effective use of mass spectrometry for glycan and glycopeptide structural analysis', *Analytical Chemistry*, 84, pp. 3040–3048. doi: 10.1021/ac3000573.

Li, X. *et al.* (2017) 'Structural Analysis of the Glycosylated Intact HIV-1 gp120-b12 Antibody Complex Using Hydroxyl Radical Protein Footprinting', *Biochemistry*, 56, pp. 957–970. doi: 10.1021/acs.biochem.6b00888.

Li, Y. *et al.* (2008) 'Removal of a single N-linked glycan in human immunodeficiency virus type 1 gp120 results in an enhanced ability to induce neutralizing antibody responses.', *Journal of virology*. American Society for Microbiology Journals, 82(2), pp. 638–51. doi: 10.1128/JVI.01691-07.

Liao, H.-X. *et al.* (2013) 'Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus', *Nature*. Nature Publishing Group, 496(7446), pp. 469–476. doi: 10.1038/nature12053.

Liu, J. *et al.* (2008) 'Molecular architecture of native HIV-1 gp120 trimers', *Nature*. Macmillan Publishers Limited. All rights reserved, 455(7209), pp. 109–113. doi: 10.1038/nature07159.

Liu, Q. *et al.* (2017) 'Quaternary contact in the initial interaction of CD4 with the HIV-1 envelope trimer.', *Nature structural & molecular biology*. NIH Public Access, 24(4), pp. 370–378. doi: 10.1038/nsmb.3382.

Lobsanov, Y. D. *et al.* (2002) 'Structure of Penicillium citrinum α1,2-mannosidase reveals the basis for differences in specificity of the endoplasmic reticulum and Golgi class I enzymes', *Journal of Biological Chemistry*, 277(7), pp. 5620–5630. doi: 10.1074/jbc.M110243200.

Ly, A. and Stamatatos, L. (2000) 'V2 loop glycosylation of the human immunodeficiency virus type 1 SF162 envelope facilitates interaction of this protein with CD4 and CCR5 receptors and protects the virus from neutralization by anti-V3 loop and anti-CD4 binding site antibodies.', *Journal of virology*. American Society for Microbiology (ASM), 74(15), pp. 6769–76.

Mahoney, M. W. and Jorgensen, W. L. (2000) 'A five-site model for liquid water and the reproduction of the

density anomaly by rigid, nonpolarizable potential functions', *The Journal of Chemical Physics*. AIP Publishing, 112(20), p. 8910. doi: 10.1063/1.481505.

Maier, J. A. *et al.* (2015) 'ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB', *Journal of Chemical Theory and Computation*. American Chemical Society, 11(8), pp. 3696–3713. doi: 10.1021/acs.jctc.5b00255.

McCoy, L. E. *et al.* (2016) 'Holes in the Glycan Shield of the Native HIV Envelope Are a Target of Trimer-Elicited Neutralizing Antibodies', *Cell Reports*, 16(9), pp. 2327–2338. doi: 10.1016/j.celrep.2016.07.074.

McCutchan, F. E. (2006) 'Global Epidemiology of HIV', *Journal of Medical Virology*, 78, pp. S7–S12. doi: 10.1002/jmv.

McKinney, W. (2010) 'Data Structures for Statistical Computing in Python', in *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, pp. 51–56.

McMichael, A. J. *et al.* (2010) 'The immune response during acute HIV-1 infection: clues for vaccine development.', *Nature reviews. Immunology*. NIH Public Access, 10(1), pp. 11–23. doi: 10.1038/nri2674.

Melo, F., Sánchez, R. and Sali, A. (2002) 'Statistical potentials for fold assessment', *Protein Science*, 11, pp. 430–448. doi: 10.1002/pro.110430.

Montefiori, D. C. (2009) 'Measuring HIV neutralization in a luciferase reporter gene assay.', *Methods in molecular biology*, 485, pp. 395–405. doi: 10.1007/978-1-59745-170-3_26.

Moremen, K. W., Tiemeyer, M. and Nairn, A. V (2012) 'Vertebrate protein glycosylation: diversity, synthesis and function.', *Nature reviews. Molecular cell biology*, 13(7), pp. 448–62. doi: 10.1038/nrm3383.

Moyo, T. *et al.* (2017) 'Chinks in the armor of the HIV-1 Envelope glycan shield: Implications for immune escape from anti-glycan broadly neutralizing antibodies', *Virology*, 501, pp. 12–24. doi: 10.1016/j.virol.2016.10.026.

O'Rourke, S. M. *et al.* (2012) 'Sequences in glycoprotein gp41, the CD4 binding site, and the V2 domain regulate sensitivity and resistance of HIV-1 to broadly neutralizing antibodies.', *Journal of virology*. American Society for Microbiology (ASM), 86(22), pp. 12105–14. doi: 10.1128/JVI.01352-12.

Pancera, M. *et al.* (2014) 'Structure and immune recognition of trimeric pre-fusion HIV-1 Env.', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited, 514(7523), pp. 455–61. doi: 10.1038/nature13808.

Panico, M. *et al.* (2016) 'Mapping the complete glycoproteome of virion-derived HIV-1 gp120 provides insights into broadly neutralizing antibody binding', *Scientific Reports*, 6(1), p. 32956. doi: 10.1038/srep32956.

Petrescu, A.-J. *et al.* (2004) 'Statistical analysis of the protein environment of N-glycosylation sites:

implications for occupancy, structure, and folding.', *Glycobiology*. Oxford University Press, 14(2), pp. 103–14. doi: 10.1093/glycob/cwh008.

Pettersen, E. F. *et al.* (2004) 'UCSF Chimera--a visualization system for exploratory research and analysis.', *Journal of computational chemistry*, 25(13), pp. 1605–12. doi: 10.1002/jcc.20084.

Plantier, J.-C. *et al.* (2009) 'A new human immunodeficiency virus derived from gorillas', *Nature Medicine*. Nature Publishing Group, 15(8), pp. 871–872. doi: 10.1038/nm.2016.

Pritchard, L. K., Harvey, D. J., *et al.* (2015) 'Cell- and Protein-Directed Glycosylation of Native Cleaved HIV-1 Envelope.', *Journal of virology*. American Society for Microbiology, 89(17), pp. 8932–44. doi: 10.1128/JVI.01190-15.

Pritchard, L. K., Vasiljevic, S., *et al.* (2015) 'Structural Constraints Determine the Glycosylation of HIV-1 Envelope Trimers.', *Cell reports*, 11(10), pp. 1604–13. doi: 10.1016/j.celrep.2015.05.017.

Rerks-Ngarm, S. *et al.* (2009) 'Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand.', *The New England journal of medicine*. Massachusetts Medical Society, 361(23), pp. 2209–20. doi: 10.1056/NEJMoa0908492.

Richman, D. D. *et al.* (2003) 'Rapid evolution of the neutralizing antibody response to HIV type 1 infection.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), pp. 4144–9. doi: 10.1073/pnas.0630530100.

Roberts, J. D., Bebenek, K. and Kunkel, T. A. (1988) 'The Accuracy of Reverse Transcriptase from HIV-1', *Science*, 242(4882), pp. 1171–1173.

Roe, D. R. and Cheatham, T. E. (2013) 'PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data', *Journal of Chemical Theory and Computation*. American Chemical Society, 9(7), pp. 3084–3095. doi: 10.1021/ct400341p.

Sagar, M. *et al.* (2006) 'Human immunodeficiency virus type 1 V1-V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity.', *Journal of virology*. American Society for Microbiology, 80(19), pp. 9586–98. doi: 10.1128/JVI.00141-06.

Šali, A. and Blundell, T. L. (1993) 'Comparative Protein Modelling by Satisfaction of Spatial Restraints', *Journal of Molecular Biology*, 234(3), pp. 779–815. doi: 10.1006/jmbi.1993.1626.

Sanders, R. W. *et al.* (2002) 'Stabilization of the Soluble, Cleaved, Trimeric Form of the Envelope Glycoprotein Complex of Human Immunodeficiency Virus Type 1', *Journal of virology*, 76(17), pp. 8875–8889. doi: 10.1128/JVI.76.17.8875.

Sanders, R. W. *et al.* (2013) 'A Next-Generation Cleaved, Soluble HIV-1 Env Trimer, BG505 SOSIP.664

gp140, Expresses Multiple Epitopes for Broadly Neutralizing but Not Non-Neutralizing Antibodies', *PLoS Pathogens*, 9(9). doi: 10.1371/journal.ppat.1003618.

Sanders, R. W. *et al.* (2016) 'HIV Neutralizing Antibodies Induced by Native-like Envelope Trimers', *Science*, 349(6244), p. aac4223. doi: 10.1126/science.aac4223.

Sather, D. N. *et al.* (2009) 'Factors associated with the development of cross-reactive neutralizing antibodies during human immunodeficiency virus type 1 infection.', *Journal of virology*, 83(2), pp. 757–69. doi: 10.1128/JVI.02036-08.

Scheid, J. F. *et al.* (2009) 'Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals', *Nature*. Nature Publishing Group, 458(7238), pp. 636–640. doi: 10.1038/nature07930.

Schiffner, T. *et al.* (2018) 'Structural and immunologic correlates of chemically stabilized HIV-1 envelope glycoproteins', *PLOS Pathogens*, 14(5), p. e1006986.

Sharp, P. and Hahn, B. H. (2011) 'Origins of HIV and the AIDS epidemic', *Cold Spring Harbor perspectives in medicine*, 1, p. a006841. doi: 10.1101/cshperspect.a006841\ra006841 [pii].

Shen, M. and Sali, A. (2006) 'Statistical potential for assessment and prediction of protein structures', *Protein Science*, 15, pp. 2507–2524. doi: 10.1110/ps.062416606.

Shrivastava, T. *et al.* (2018) 'Envelope proteins of two HIV-1 clades induced different epitope-specific antibody response', *Vaccine*. Elsevier Ltd, pp. 1–10. doi: 10.1016/j.vaccine.2018.01.081.

Siliciano, R. F. and Greene, W. C. (2011) 'HIV latency.', *Cold Spring Harbor perspectives in medicine*, 1(1), p. a007096. doi: 10.1101/cshperspect.a007096.

Simek, M. D. *et al.* (2009) 'Human Immunodeficiency Virus Type 1 Elite Neutralizers: Individuals with Broad and Potent Neutralizing Activity Identified by Using a High-Throughput Neutralization Assay together with an Analytical Selection Algorithm', *Journal of Virology*. American Society for Microbiology, 83(14), pp. 7337–7348. doi: 10.1128/JVI.00110-09.

Simpson, R. J. (Cold S. H. L. (2003) *Proteins and Proteomics: A Laboratory Manual*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Stambulchik, E. (1998) *Grace*. Available at: http://plasma-gate.weizmann.ac.il/Grace/.

Stanley, P., Taniguchi, N. and Aebi, M. (2017) 'N-Glycans', in *Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology [Internet]*. 3rd edn. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. doi: 10.1101/glycobiology.3e.009.

Starcich, B. R. *et al.* (1986) 'Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS', *Cell*, 45(5), pp. 637–648. doi: 10.1016/0092-

8674(86)90778-6.

Stephenson, K. E. and Barouch, D. H. (2016) 'Broadly Neutralizing Antibodies for HIV Eradication.', *Current HIV/AIDS reports*. doi: 10.1007/s11904-016-0299-7.

Stewart-Jones, G. B. E. *et al.* (2016) 'Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G.', *Cell*. Elsevier, 165(4), pp. 813–826. doi: 10.1016/j.cell.2016.04.010.

Suga, A., Nagae, M. and Yamaguchi, Y. (2018) 'Analysis of protein landscapes around N-glycosylation sites from the PDB repository for understanding the structural basis of N-glycoprotein processing and maturation', *Glycobiology*, 28(10), pp. 774–785. doi: 10.1093/glycob/cwy059.

Summers, M. E. *et al.* (1992) 'Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1', *Protein Science*, pp. 563–574.

Travers, S. A. (2012) 'Conservation, Compensation, and Evolution of N-Linked Glycans in the HIV-1 Group M Subtypes and Circulating Recombinant Forms', *ISRN AIDS*, 2012, pp. 1–9. doi: 10.5402/2012/823605.

Turner, P. J. (1991) *Xmgr*. Center for Coastal and Land-Margin Research Oregon Graduate Institute of Science and Technology Beaverton, Oregon. Available at: http://plasma-gate.weizmann.ac.il/Xmgr/.

Upadhyay, C. *et al.* (2014) 'Distinct mechanisms regulate exposure of neutralizing epitopes in the V2 and V3 loops of HIV-1 envelope.', *Journal of virology*. American Society for Microbiology Journals, 88(21), pp. 12853–65. doi: 10.1128/JVI.02125-14.

Utachee, P. *et al.* (2010) 'Two N-linked glycosylation sites in the V2 and C2 regions of human immunodeficiency virus type 1 CRF01_AE envelope glycoprotein gp120 regulate viral neutralization susceptibility to the human monoclonal antibody specific for the CD4 binding domain.', *Journal of virology*. American Society for Microbiology Journals, 84(9), pp. 4311–20. doi: 10.1128/JVI.02619-09.

Varki, A. *et al.* (2009) *Essentials of Glycobiology*. 2nd edn. New York: Cold Spring Harbor Laboratory Press.

Wagh, K. *et al.* (2018) 'Completeness of HIV-1 Envelope Glycan Shield at Transmission Determines Neutralization Breadth', *Cell Reports*, 25(4), pp. 893-908.e7. doi: 10.1016/j.celrep.2018.09.087.

Walker, L. M. *et al.* (2009) 'Broad and Potent Neutralizing Antibodies from an African Donor Reveal a New HIV-1 Vaccine Target', *Science*, 326(5950), pp. 285–289. doi: 10.1126/science.1178746.

Walker, L. M. *et al.* (2010) 'A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals.', *PLoS pathogens*. Public Library of Science, 6(8), p. e1001028. doi: 10.1371/journal.ppat.1001028.

Walker, L. M. *et al.* (2011) 'Broad neutralization coverage of HIV by multiple highly potent antibodies.', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 477(7365),

pp. 466–70. doi: 10.1038/nature10373.

Ward, A. B. and Wilson, I. A. (2015) 'Insights into the trimeric HIV-1 envelope glycoprotein structure', *Trends in Biochemical Sciences*, 40(2), pp. 101–107. doi: 10.1016/j.tibs.2014.12.006.

Wei, X. *et al.* (2003) 'Antibody neutralization and escape by HIV-1', *Nature*, 422(6929), pp. 307–312. doi: 10.1038/nature01470.

WHO (2016) *WHO | HIV/AIDS*. World Health Organization. Available at: http://www.who.int/mediacentre/factsheets/fs360/en/ (Accessed: 31 May 2017).

WHO (2017) *WHO | HIV/AIDS Factsheet 2017*. Available at: http://www.who.int/mediacentre/factsheets/fs360/en/.

Willey, R. L. *et al.* (1986) 'Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus.', *Proceedings of the National Academy of Sciences*, 83(14), pp. 5038–5042. doi: 10.1073/pnas.83.14.5038.

Wood, N. T. *et al.* (2013) 'The influence of N-linked glycans on the molecular dynamics of the HIV-1 gp120 V3 loop.', *PLoS ONE*. Public Library of Science, 8(11), p. e80301. doi: 10.1371/journal.pone.0080301.

Woods Group (2005) *GLYCAM-Web, Complex Carbohydrate Research Center, University of Georgia, Athens, GA.* Available at: http://glycam.org.

Woods, R. J. and Tessier, M. B. (2010) 'Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes.', *Current opinion in structural biology*, 20(5), pp. 575–83. doi: 10.1016/j.sbi.2010.07.005.

Wu, X. *et al.* (2010) 'Rational Design of Envelope Identifies Broadly NeutralizingHuman Monoclonal Antibodies to HIV-1', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 329(5993), pp. 853–856. doi: 10.1126/science.1188321.

Wu, X. *et al.* (2011) 'Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing', *Science*, 333(6049), pp. 1593–1602. doi: 10.1126/science.1207532.

Wyatt, R. *et al.* (1995) 'Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding.', *Journal of virology*. American Society for Microbiology Journals, 69(9), pp. 5723–33. doi: 10.1128/jvi.75.7.3435-3443.2001.

Wyatt, R. *et al.* (1998) 'The antigenic structure of the HIV gp120 envelope glycoprotein', *Nature*. Nature Publishing Group, 393(6686), pp. 705–711. doi: 10.1038/31514.

Wyatt, R. and Sodroski, J. (1998) 'The HIV-1 Envelope Glycoproteins: Fusogens,Antigens, and Immunogens', *Science*. American Association for the Advancement of Science, 280(5371), pp. 1884–8. doi:

10.1126/science.280.5371.1884.

Xiang, Y., Karaveg, K. and Moremen, K. W. (2016) 'Substrate recognition and catalysis by GH47 α-mannosidases involved in Asn-linked glycan maturation in the mammalian secretory pathway.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(49), pp. E7890–E7899. doi: 10.1073/pnas.1611213113.

Yang, M. *et al.* (2017) 'Conformational Heterogeneity of the HIV Envelope Glycan Shield', *Scientific Reports*. Springer US, 7(1), pp. 1–15. doi: 10.1038/s41598-017-04532-9.

Yokoyama, M. *et al.* (2016) 'In silico Analysis of HIV-1 Env-gp120 Reveals Structural Bases for Viral Adaptation in Growth-Restrictive Cells.', *Frontiers in microbiology*. Frontiers Media SA, 7, p. 110. doi: 10.3389/fmicb.2016.00110.

Yu, W.-H. *et al.* (2018) 'Exploiting glycan topography for computational design of Env glycoprotein antigenicity', *PLOS Computational Biology*. Edited by G. Tucker-Kellogg. Public Library of Science, 14(4), p. e1006093. doi: 10.1371/journal.pcbi.1006093.

Zhou, T. *et al.* (2007) 'Structural definition of a conserved neutralization epitope on HIV-1 gp120', *Nature*. Nature Publishing Group, 445(7129), pp. 732–737. doi: 10.1038/nature05580.

Zhou, T. *et al.* (2010) 'Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01', *Science*, 329(5993), pp. 811–817. doi: 10.1126/science.1192819.

Zhou, T. *et al.* (2017) 'Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation Article', *CellReports*. ElsevierCompany., 19(4), pp. 719–732. doi: 10.1016/j.celrep.2017.04.013.